

APPLICATION
FOR
UNITED STATES LETTERS PATENT

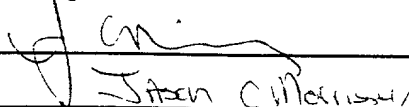
TITLE: LEPTIN INDUCED GENES

APPLICANT: DAVID WHITE, JIANGHONG ZHOU, AND LOUIS A. TARTAGLIA

"EXPRESS MAIL" Mailing Label Number EL245413311 US

Date of Deposit November 19 1998

I hereby certify under 37 CFR 1.10 that this correspondence is being deposited with the United States Postal Service as "Express Mail Post Office To Addressee" with sufficient postage on the date indicated above and is addressed to the Assistant Commissioner for Patents, Washington, D.C. 20231.


Jason C. McCreary

LEPTIN INDUCED GENES5 Background of the Invention

The *ob* gene product, leptin, is an important circulating regulator of body weight. Leptin binds to and activates the long form of ObR, the leptin receptor (Tartaglia et al. (1995) *Cell* 83:1263-71). Leptin is
10 thought to modulate body weight by influencing appetite and other factors. Compounds other than leptin, e.g., neuropeptide Y, melanocortins, CART, and orexins are also thought to play a role in modulation of body weight by influencing factors such as appetite and satiety, fat
15 storage, and energy output.

Summary of the Invention

The present invention is based, at least in part, on the identification of six genes whose expression is induced
20 by leptin. Two of these genes, LIG46 and LIG56, are novel genes. Four of the genes have been previously identified. The previously identified genes are: Tgtp, encoding a T cell-specific GTP-binding protein; LRG-47, encoding an interferon (IFN) inducible protein; RC10-II, encoding a
25 subunit of a 20S brain proteasome; and Stra13, encoding a retinoic acid inducible protein.

The six leptin-inducible genes of the invention and the proteins they encode represent targets for the development of therapeutic agents for use in modulating body
30 weight. For example, agents that alter the expression or activity of one or more of the leptin-induced proteins can be used to modulate body weight. Such agents can be identified using cellular, *in vitro*, or *in vivo* assays which

monitor the expression or activity of one or more of the six
leptin-induced proteins. Potentially useful therapeutic
agents can also be identified through the use of assays
designed to identify agents that bind to one of the leptin-
5 induced proteins. The leptin-induced genes of the invention
and the proteins they encode may themselves be useful
therapeutically and diagnostically.

LIG46

10 The murine LIG46 cDNA described below (SEQ ID NO:1)
has a 1191 nucleotide open reading frame (nucleotides 3 -
1193 of SEQ ID NO:1; SEQ ID NO:3) which encodes a 397 amino
acid protein (SEQ ID NO:2). This protein includes a
predicted signal sequence of about 32 amino acids (from
15 amino acid 1 to about amino acid 32 of SEQ ID NO:2) and a
predicted mature protein of about 365 amino acids (from
about amino acid 33 to amino acid 397 of SEQ ID NO:2; SEQ ID
NO:4). The extracellular domain of LIG46 extends from about
amino acid 33 to about amino acid 302. LIG46 possesses one
20 predicted transmembrane domain which extends from about
amino acid 303 (extracellular end) to about 320
(intracellular end) of SEQ ID NO:2. The cytoplasmic domain
of LIG46 extends from about amino acid 321 to about amino
acid 397.

25 The human LIG46 cDNA described below (SEQ ID NO:__)
has a 1191 nucleotide open reading frame which encodes a 397
amino acid protein (SEQ ID NO:__). This protein includes a
predicted signal sequence of about 32 amino acids (from
amino acid 1 to about amino acid 32 of SEQ ID NO:__) and a
30 predicted mature protein of about 365 amino acids (from
about amino acid 33 to amino acid 397 of SEQ ID NO:__; SEQ

ID NO:___).

LIG46 protein has some sequence similarity to a number of galactosyltransferases. Galactosyltransferases have been implicated in developmental processes. In addition, galactosyltransferases may play a role in cell-to-cell signaling by modifying the carbohydrate repertoire on cell surface receptors to activate, inhibit or otherwise modify (e.g., by altering receptor affinity for a ligand) receptor activity. Thus, LIG46 may play a role body weight regulation by influencing cell-to-cell signaling mediated by molecules involved in body weight regulation, e.g., leptin.

The LIG46 polypeptide sequence of SEQ ID NO:2 includes potential N-glycosylation sites at amino acids 30-33, 79-82, 89-92, 127-173, and 219-222; potential protein kinase C phosphorylation sites at amino acids 54-56, 202-204, 221-223, 323-325, and 377-379; potential casein kinase II phosphorylation sites at amino acids 31-34, 94-97, 185-188, 221-224, 234-237, and 368-371; a potential tyrosine kinase phosphorylation site at amino acids 115-122; and a potential amidation site at amino acids 3-6.

In one aspect, the invention provides isolated nucleic acid molecules encoding LIG46 proteins or biologically active portions thereof, as well as nucleic acid molecules suitable for use as primers or hybridization probes for the detection of LIG46-encoding nucleic acid molecules.

The invention further provides nucleic acid molecules that are at least 45% (or 55%, 65%, 75%, 85%, 95%, or 98%) identical to the nucleotide sequence shown in SEQ ID NO:1, or SEQ ID NO:3, or a complement thereof.

The invention provides a nucleic acid molecule which

includes a fragment of at least 300 (325, 350, 375, 400, 425, 450, 500, 550, 600, 650, 700, 800, 900, 1000, 1200, 1300, or 1400) nucleotides of the nucleotide sequence shown in SEQ ID NO:1, or SEQ ID NO:3, or a complement thereof.

5 The invention also features a nucleic acid molecule which includes a nucleotide sequence encoding a protein having an amino acid sequence that is at least 45% (or 55%, 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:4.

10 In a preferred embodiment, a LIG46 nucleic acid molecule has the nucleotide sequence shown SEQ ID NO:1 or SEQ ID NO:3.

Also within the invention is a nucleic acid molecule which encodes a fragment of a polypeptide having the amino
15 acid sequence of SEQ ID NO:2 or SEQ ID NO:4, the fragment including at least 15 (25, 30, 50, 100, 150, 300, or 390) contiguous amino acids of SEQ ID NO:2 or SEQ ID NO:4.

The invention includes a nucleic acid molecule which encodes a naturally occurring allelic variant of a
20 polypeptide comprising the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:4, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule comprising SEQ ID NO:1 or SEQ ID NO:3 under stringent conditions.

Also within the invention are: an isolated LIG46
25 protein having an amino acid sequence that is at least about 65%, preferably 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:4 (mature murine LIG46) or the amino acid sequence of SEQ ID NO:2 (immature murine LIG46); and an isolated LIG46 protein having an amino acid sequence
30 that is at least about 85%, 95%, or 98% identical to a portion of LIG46 having homology to a galactosyltransferase

(e.g., amino acids 192-353, 142-184, 201-296, 289-347, 140-183, 367-391, 177-266, 299-343, or 140-184 of SEQ ID NO:2) or a neurogenic secreted signalling protein (e.g., amino acids 200-291, 270-354, 144-183, 380-394, or 211-248 of SEQ ID NO:2).

Also within the invention are: an isolated LIG46 protein which is encoded by a nucleic acid molecule having a nucleotide sequence that is at least about 65%, preferably 75%, 85%, or 95% identical to SEQ ID NO:3; and an isolated LIG46 protein which is encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule having the nucleotide sequence of the complement of SEQ ID NO:3.

Also within the invention is a polypeptide which is a naturally occurring allelic variant of a polypeptide that includes the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:4, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising the complement of SEQ ID NO:1 or SEQ ID NO:3 under stringent conditions.

Another embodiment of the invention provides LIG46 nucleic acid molecules which specifically detect LIG46 nucleic acid molecules (e.g., a nucleic acid molecule encoding human LIG46) relative to nucleic acid molecules encoding other galactosyltransferases. For example, in one embodiment, a LIG46 nucleic acid molecule hybridizes under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, or a complement thereof, but does not hybridize to unrelated galactosyltransferases. In another embodiment, the LIG46

nucleic acid molecule is at least 300 (325, 350, 375, 400, 425, 450, 500, 550, 600, 650, 700, 800, 900, 1000, or 1200) nucleotides in length and hybridizes under stringent conditions to a nucleic acid molecule comprising the
5 nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, or a complement thereof.

Another aspect of the invention provides a vector, e.g., a recombinant expression vector, comprising a LIG46 nucleic acid molecule of the invention. In another
10 embodiment the invention provides a host cell containing such a vector. The invention also provides a method for producing LIG46 protein by culturing, in a suitable medium, a host cell of the invention containing a recombinant expression vector such that a LIG46 protein is produced.

15 Another aspect of this invention provides isolated or recombinant LIG46 proteins and polypeptides. Preferred LIG46 proteins and polypeptides possess at least one biological activity possessed by naturally-occurring LIG46 (e.g., the ability to act as a galactosyl transferase) and
20 are induced by leptin.

The LIG46 proteins of the present invention, or biologically active portions thereof, can be operatively linked to a non-LIG46 polypeptide (e.g., heterologous amino acid sequences) to form LIG46 fusion proteins. The
25 invention further features antibodies that specifically bind LIG46 proteins, such as monoclonal or polyclonal antibodies.

In addition, the LIG46 proteins or biologically active portions thereof can be incorporated into pharmaceutical compositions, which optionally include pharmaceutically
30 acceptable carriers.

In another aspect, the present invention provides a

method for detecting the presence of LIG46 activity or expression in a biological sample by contacting the biological sample with an agent capable of detecting an indicator of LIG46 activity such that the presence of LIG46 activity is detected in the biological sample.

In another aspect, the invention provides a method for modulating LIG46 activity comprising contacting a cell with an agent that modulates (inhibits or stimulates) LIG46 activity or expression such that LIG46 activity or expression in the cell is modulated. In one embodiment, the agent is an antibody that specifically binds to LIG46 protein. In another embodiment, the agent modulates expression of LIG46 by modulating transcription of a LIG46 gene, splicing of a LIG46 mRNA, or translation of a LIG46 mRNA. In yet another embodiment, the agent is a nucleic acid molecule having a nucleotide sequence that is antisense to the coding strand of the LIG46 mRNA or the LIG46 gene.

In one embodiment, the methods of the present invention are used to treat a subject having a disorder characterized by and undesirable level of LIG46 protein or nucleic acid expression or activity by administering an agent which is a LIG46 modulator to the subject. In one embodiment, the LIG46 modulator is a LIG46 protein. In another embodiment the LIG46 modulator is a LIG46 nucleic acid molecule. In other embodiments, the LIG46 modulator is a peptide, peptidomimetic, or other small molecule. In a preferred embodiment, the disorder is obesity or cachexia.

For treatment of obesity it is desirable to administer an agent which reduces the expression or activity of LIG46 (an LIG46 antagonist). Such an agent can be administered in conjunction with leptin. Preferably the

amount of leptin administered is sufficient, in combination with any endogenous leptin, to render the subject being treated sensitive to the effects of the LIG46 antagonist.

For treatment of low body weight it is desirable to
5 administer an agent which increases the expression of activity of LIG46 (an LIG46 agonist).

The present invention also provides a diagnostic assay for identifying the presence or absence of a genetic lesion or mutation characterized by at least one of: (i)
10 aberrant modification or mutation of a gene encoding a LIG46 protein; (ii) mis-regulation of a gene encoding a LIG46 protein; and (iii) aberrant post-translational modification of a LIG46 protein, wherein a wild-type form of the gene encodes a protein with a LIG46 activity.

15 In another aspect, the invention provides a method for identifying a compound that binds to or modulates the activity of a LIG46 protein. In general, such methods entail measuring a biological activity of a LIG46 protein in the presence and absence of a test compound and identifying
20 those compounds which alter the activity of the LIG46 protein.

The invention also features methods for identifying a compound which modulates the expression of LIG46 by measuring the expression of LIG46 in the presence and
25 absence of a compound.

LIG56

The murine LIG56 cDNA described below (SEQ ID NO:5) has a 1200 nucleotide open reading frame (nucleotides 1 -
30 1200 of SEQ ID NO:5; SEQ ID NO:7) which encodes a 400 amino acid protein (SEQ ID NO:6).

The LIG56 polypeptide sequence of SEQ ID NO:6 includes potential N-glycosylation sites at amino acids 252-255; potential protein kinase C phosphorylation sites at amino acids 67-69, 75-77, 203-205, 218-220, 295-297, and 5 299-301; potential casein kinase II phosphorylation sites at amino acids 126-129, 170-173, 203-206, 256-259, 291-294, 341-344, and 345-349; a potential tyrosine kinase phosphorylation site at amino acids 233-241; potential N-myristylation sites at amino acids 66-71, 85-90, 116-121, and 10 308-313; and a potential amidation site at amino acids 63-70.

LIG56 may be a GTP-binding protein. Portions of LIG56 protein are too similar to one or more murine GTP-binding proteins (Genbank Accession Numbers: L38444; U15636; 15 M63630; U19119; and U53219).

LIG56 protein possesses a GTP-binding protein-like domain (amino acids 12 to 283 of SEQ ID NO:6) and an LRG-47-like domain (amino acids 24-177 of SEQ ID NO:6).

In one aspect, this invention provides isolated 20 nucleic acid molecules encoding LIG56 proteins or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection of LIG56-encoding nucleic acids.

The invention provides a nucleic acid molecule which 25 is at least 45% (or 55%, 65%, 75%, 85%, 95%, or 98%) identical to the nucleotide sequence shown in SEQ ID NO:5 or SEQ ID NO:7, or a complement thereof.

The invention provides a nucleic acid molecule which includes a fragment of at least 100 (200, 250, 300, 350, 30 400, 450, 500, 550, 600, 650, 700, 800, 900, 1000, 1100, or 1200) nucleotides of the nucleotide sequence shown in SEQ ID

NO:5 or SEQ ID NO:7, or a complement thereof.

The invention also features a nucleic acid molecule which includes a nucleotide sequence encoding a protein having an amino acid sequence that is at least 45% (or 55%, 5 65%, 75%, 85%, 95%, or 98%) identical to the amino acid sequence of SEQ ID NO:6.

In a preferred embodiment, a LIG56 nucleic acid molecule has the nucleotide sequence shown SEQ ID NO:5 or SEQ ID NO:7.

10 Also within the invention is a nucleic acid molecule which encodes a fragment of a polypeptide having the amino acid sequence of SEQ ID NO:6, the fragment including at least 15 (25, 30, 50, 100, 150, 300, or 400) contiguous amino acids of SEQ ID NO:6.

15 The invention includes a nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:6, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule having the sequence of the complement 20 of SEQ ID NO:5 or SEQ ID NO:7 under stringent conditions.

Also within the invention are: an isolated LIG56 protein having an amino acid sequence that is at least about 65%, preferably 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:6.

25 Also within the invention are: an isolated LIG56 protein which is encoded by a nucleic acid molecule having a nucleotide sequence that is at least about 65%, preferably 75%, 85%, or 95% identical to SEQ ID NO:7; and an isolated LIG56 protein which is encoded by a nucleic acid molecule 30 having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid

molecule having the nucleotide sequence of SEQ ID NO:7.

Also within the invention is a polypeptide which is a naturally occurring allelic variant of a polypeptide that includes the amino acid sequence of SEQ ID NO:6, wherein the
5 polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO:5 or SEQ ID NO:7 under stringent conditions.

Another embodiment of the invention provides LIG56 nucleic acid molecules which specifically detect LIG56
10 nucleic acid molecules (e.g., human LIG56) relative to nucleic acid molecules encoding other unrelated nucleic acid molecules having sequence homology to GTP-binding proteins.

For example, in one embodiment, a LIG56 nucleic acid molecule hybridizes under stringent conditions to a nucleic
15 acid molecule comprising the nucleotide sequence of SEQ ID NO:5 or SEQ ID NO:7, or a complement thereof. In another embodiment, the LIG56 nucleic acid molecule is at least 300 (325, 350, 375, 400, 425, 450, 500, 550, 600, 650, 700, 800, 900, 1000, 1100 or 1200) nucleotides in length and
20 hybridizes under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence shown in SEQ ID NO:5 or SEQ ID NO:7, or a complement thereof. In another embodiment, the invention provides an isolated nucleic acid molecule which is antisense to the coding strand of a LIG56
25 nucleic acid.

Another aspect of the invention provides a vector, e.g., a recombinant expression vector, comprising a LIG56 nucleic acid molecule of the invention. In another embodiment the invention provides a host cell containing
30 such a vector. The invention also provides a method for producing LIG56 protein by culturing, in a suitable medium,

a host cell of the invention containing a recombinant expression vector such that a LIG56 polypeptide is produced.

Another aspect of this invention provides isolated or recombinant LIG56 proteins and polypeptides. Preferred
5 LIG56 proteins and polypeptides possess at least one biological activity possessed by naturally occurring LIG56 and are induced by leptin.

The LIG56 proteins of the present invention, or biologically active portions thereof, can be operatively
10 linked to a non-LIG56 polypeptide (e.g., heterologous amino acid sequences) to form LIG56 fusion proteins. The invention further features antibodies that specifically bind LIG56 proteins, such as monoclonal or polyclonal antibodies.

In addition, the LIG56 proteins or biologically active
15 portions thereof can be incorporated into pharmaceutical compositions, which optionally include pharmaceutically acceptable carriers.

In another aspect, the present invention provides a method for detecting the presence of LIG56 activity or
20 expression in a biological sample by contacting the biological sample with an agent capable of detecting an indicator of LIG56 activity such that the presence of LIG56 activity is detected in the biological sample.

In another aspect, the invention provides a method
25 for modulating LIG56 activity comprising contacting a cell with an agent that modulates (inhibits or stimulates) LIG56 activity or expression such that LIG56 activity or expression in the cell is modulated. In one embodiment, the agent is an antibody that specifically binds to LIG56
30 protein. In another embodiment, the agent modulates expression of LIG56 by modulating transcription of a LIG56

gene, splicing of a LIG56 mRNA, or translation of a LIG56 mRNA. In yet another embodiment, the agent is a nucleic acid molecule having a nucleotide sequence that is antisense to the coding strand of the LIG56 mRNA or the LIG56 gene.

5 In one embodiment, the methods of the present invention are used to treat a subject having a disorder characterized by an undesirable level of LIG56 protein or nucleic acid expression (e.g., a body weight disorder) or activity by administering an agent which is a LIG56
10 modulator to the subject. In one embodiment, the LIG56 modulator is a LIG56 protein. In another embodiment the LIG56 modulator is a LIG56 nucleic acid molecule. In other embodiments, the LIG56 modulator is a peptide, peptidomimetic, or other small molecule. In a preferred
15 embodiment, the disorder is obesity or cachexia.

The present invention also provides a diagnostic assay for identifying the presence or absence of a genetic lesion or mutation characterized by at least one of: (i) aberrant modification or mutation of a gene encoding a LIG56
20 protein; (ii) mis-regulation of a gene encoding a LIG56 protein; and (iii) aberrant post-translational modification of a LIG56 protein, wherein a wild-type form of the gene encodes a protein with a LIG56 activity.

In another aspect, the invention provides a method
25 for identifying a compound that binds to or modulates the activity of a LIG56 protein. In general, such methods entail measuring a biological activity of a LIG56 protein in the presence and absence of a test compound and identifying those compounds which alter the activity of the LIG56
30 protein.

The invention also features methods for identifying

a compound which modulates the expression of LIG56 by measuring the expression of LIG56 in the presence and absence of a compound.

5 Tgtp, LRG-47, RC10-II, and Stra13

Tgtp, LRG-47, RC10-II, and Stra13 are known genes. However, none of these genes has previously been implicated in body weight regulation. The present invention is based, in part, on the discovery that expression of each of these
10 genes is induced by leptin. Because Tgtp, LRG-47, RC10-II, and Stra13 are induced by leptin, Tgtp, LRG-47, RC10-II, and Stra13 protein and the nucleic acid molecules encoding them are useful in the development of therapeutic compounds for the treatment or prevention of body weight disorders.

15 Tgtp (Genbank Accession Number L38444) encodes a T cell-specific guanine nucleotide triphosphate-binding protein (Carlow et al. (1994) *J. Immunol.* 154:1724-34).

LRG-47 (Genbank Accession Number U19119) is induced by LPS, IFN- γ , and IFN- α/β and encodes a protein that has
20 some homology to GTP-binding proteins (Sorace et al. (1995) *J. Leukocyte Biol.* 58:477-84).

LRG-47 (Genbank Accession Number U19119) is a LPS, IFN- γ , and IFN- α/β -inducible gene having homology to IRG-47 and Mg21, both of which are IFN- γ -inducible genes (Sorace et
25 al. (1995) *J. Leukocyte Biol.* 58:477-484). LRG-47 also has homology to Tgtp and may be a GTP-binding protein (Sorace et al., *supra*).

RC10-II (Genbank Accession Number D21800) is gene that encodes the RC10-II subunit of the 20S proteasome of
30 rat embryonic brain (Nishimura et al. (1993) *FEBS Lett.* 336:462-66). It has been suggested that RC10-II is a

proteasomal subunit that is required for expression of tryptic activity (Nishimura et al., *supra*).

Stral3 (Genbank Accession Number AF010305) is a retinoic acid-inducible gene that encodes a basic helix-loop-helix protein (Boudjelal et al. (1997) *Genes Dev.* 11:2052-65). Stral3 may act as a repressor of activated transcription and is thought to play a role in neuronal differentiation (Boudjelal et al., *supra*).

The invention provides a method for identifying a compound that binds to or modulates the activity of a Tgtp, LRG-47, RC10-II, or Stral3 protein. In general, such methods entail measuring a biological activity of a Tgtp, LRG-47, RC10-II, or Stral3 protein in the presence and absence of a test compound and identifying those compounds which bind to or alter the activity of the Tgtp, LRG-47, RC10-II, or Stral3 protein.

The invention also features methods for identifying a compound which modulates the expression of Tgtp, LRG-47, RC10-II, or Stral3 by measuring the expression of Tgtp, LRG-47, RC10-II, or Stral3 in the presence and absence of a compound.

Thus, the invention provides a method for modulating Tgtp, LRG-47, RC10-II, or Stral3 activity comprising contacting a cell with an agent that modulates (inhibits or stimulates) Tgtp, LRG-47, RC10-II, or Stral3 activity or expression such that Tgtp, LRG-47, RC10-II, or Stral3 activity or expression in the cell is modulated. In one embodiment, the agent is an antibody that specifically binds to Tgtp, LRG-47, RC10-II, or Stral3 protein. In another embodiment, the agent modulates expression of Tgtp, LRG-47, RC10-II, or Stral3 by modulating transcription of a Tgtp,

LRG-47, RC10-II, or Stra13 gene; splicing of a Tgtp, LRG-47, RC10-II, or Stra13 mRNA; or translation of a Tgtp, LRG-47, RC10-II, or Stra13 mRNA. In yet another embodiment, the agent is a nucleic acid molecule having a nucleotide
5 sequence that is antisense to the coding strand of the Tgtp, LRG-47, RC10-II, or Stra13 mRNA or the Tgtp, LRG-47, RC10-II, or Stra13 gene.

In one embodiment, the methods of the present invention are used to treat a subject having a disorder
10 influenced by Tgtp, LRG-47, RC10-II, or Stra13 protein or nucleic acid expression or activity by administering an agent which is a Tgtp, LRG-47, RC10-II, or Stra13 modulator to the subject. In one embodiment, the Tgtp, LRG-47, RC10-II, or Stra13 modulator is a Tgtp, LRG-47, RC10-II, or
15 Stra13 protein. In another embodiment the Tgtp, LRG-47, RC10-II, or Stra13 modulator is a Tgtp, LRG-47, RC10-II, or Stra13 nucleic acid molecule. In other embodiments, the Tgtp, LRG-47, RC10-II, or Stra13 modulator is a peptide, peptidomimetic, or other small molecule. In a preferred
20 embodiment, the disorder is obesity or cachexia.

The present invention also provides a diagnostic assay for identifying the presence or absence of a genetic lesion or mutation characterized by at least one of: (i) aberrant modification or mutation of a gene encoding a Tgtp,
25 LRG-47, RC10-II, or Stra13 protein; (ii) mis-regulation of a gene encoding a Tgtp, LRG-47, RC10-II, or Stra13 protein; and (iii) aberrant post-translational modification of a Tgtp, LRG-47, RC10-II, or Stra13 protein, wherein a wild-type form of the gene encodes a protein with a Tgtp, LRG-47,
30 RC10-II, or Stra13 activity, as such lesion are characterized by body weight disorders.

In another aspect, the present invention provides a method for detecting the presence of Tgtp, LRG-47, RC10-II, or Stral3 activity or expression in a biological sample by contacting the biological sample with an agent capable of
5 detecting an indicator of Tgtp, LRG-47, RC10-II, or Stral3 activity such that the presence of Tgtp, LRG-47, RC10-II, or Stral3 activity is detected in the biological sample.

Other features and advantages of the invention will be apparent from the following detailed description and
10 claims.

Brief Description of the Drawings

Figure 1 depicts the cDNA sequence (SEQ ID NO:1) and predicted amino acid sequence (SEQ ID NO:2) of murine LIG46.

15 Figure 2 depicts a series of alignments of the amino acid sequence of LIG46 with portions of a number of galactosyltransferases, including (from top to bottom): *Mus musculus* UDP-Gal: betaGlcNAc beta 1,3-galactosyltransferase-I (Accession Number AF029790; SEQ ID NO:__); *Mus musculus*
20 IPP-Gal: betaGlcNAc beta 1,3-galactosyltransferase-III (Accession Number AF029792); *Drosophila melanogaster* neurogenic secreted signalling protein ("Brainiac"; Accession Number U41449; SEQ ID NO:__); and *Homo sapiens* UDP-galactose: 2-acetamido-2-deoxy-D-glucose3beta-
25 galactosyltransferase (Accession Number Y15014; SEQ ID NO:__). The amino acid sequence above the solid line is a majority sequence

Figure 3 is a hydropathy plot of LIG46. The location of the predicted transmembrane (TM), cytoplasmic
30 (IN), and extracellular (OUT) domains are indicated as are the position of cysteines (cys; vertical bars immediately

below the plot). Relative hydrophobicity is shown above the dotted line, and relative hydrophilicity is shown below the dotted line.

Figure 4 depicts the cDNA sequence (SEQ ID NO:5) and
5 predicted amino acid sequence (SEQ ID NO:6) of murine LIG56.

Figure 5 is a hydropathy plot of LIG56. Relative hydrophobicity is shown above the dotted line, and relative hydrophilicity is shown below the dotted line.

Figure 6 is a graph depicting the effect of LIG46
10 sense and antisense oligonucleotides on food intake of male obese (ob/ob) mice in the presence and absence of leptin.

Figure 7 depicts the cDNA sequence of human LIG46

Figure 8 depicts the predicted amino acid sequence
of human LIG46.

Figure 9 depicts an alignment of the cDNA sequences
15 of human LIG46 (upper sequence) and murine LIG46 (lower sequence).

Figure 10 depicts an alignment of the predicted
amino acid sequences of human LIG46 (upper sequence) and
20 murine LIG46 (lower sequence).

Figure 11 is a graph depicting the effect of LIG46
sense and antisense oligonucleotides on food intake of male lean mice in the presence and absence of leptin.

25 Detailed Description of the Invention

The present invention is based, in part, on the identification of six genes whose expression is induced by leptin. Four of the genes, Tgtp, LRG-47, RC10-II, and Stra13, are known genes. Two of the genes, LIG46 and LIG56,
30 are novel.

A nucleotide sequence encoding murine LIG46 protein

is shown in Figure 1 (SEQ ID NO:1; SEQ ID NO:3 includes the open reading frame only). A predicted amino acid sequence of LIG46 protein is also shown in Figure 1 (SEQ ID NO: 2).

The murine LIG46 cDNA of Figure 1 (SEQ ID NO:1)
5 encodes a 397 amino acid protein.

Murine LIG46 is one member of a family of molecules (the "LIG46 family") having certain conserved structural and functional features. The term "family" when referring to the protein and nucleic acid molecules of the invention is
10 intended to mean two or more proteins or nucleic acid molecules having a common structural domain and having sufficient amino acid or nucleotide sequence identity as defined herein. Such family members can be naturally occurring and can be from either the same or different
15 species. For example, a family can contain a first protein of murine origin and a homologue of that protein of human origin, as well as a second, distinct protein of human origin and a murine homologue of that protein. Members of a family may also have common functional characteristics.

20 A nucleotide sequence encoding murine LIG56 protein is shown in Figure 4 (SEQ ID NO:5; SEQ ID NO:7 includes the open reading frame only). A predicted amino acid sequence of LIG46 protein is also shown in Figure 4 (SEQ ID NO:6).

The murine LIG56 cDNA of Figure 4 (SEQ ID NO:5)
25 encodes a 400 amino acid protein.

Murine LIG56 is one member of a family of molecules (the "LIG56 family") having certain conserved structural and functional features. The term "family" when referring to the protein and nucleic acid molecules of the invention is
30 intended to mean two or more proteins or nucleic acid molecules having a common structural domain and having

sufficient amino acid or nucleotide sequence identity as defined herein. Such family members can be naturally occurring and can be from either the same or different species. For example, a family can contain a first protein
5 of murine origin and a homologue of that protein of human origin, as well as a second, distinct protein of human origin and a murine homologue of that protein. Members of a family may also have common functional characteristics.

Tgtp (Genbank Accession Number L38444) encodes a T
10 cell-specific guanine nucleotide triphosphate-binding protein (Carlow et al. (1994) *J. Immunol.* 154:1724-34).

LRG-47 (Genbank Accession Number U19119) is induced by LPS, IFN- γ , and IFN- α/β and encodes a protein that has some homology to GTP-binding proteins (Sorace et al. (1995)
15 *J. Leukocyte Biol.* 58:477-84).

LRG-47 (Genbank Accession Number U19119) is a LPS, IFN- γ , and IFN- α/β -inducible gene having homology to IRG-47 and Mg21, both of which are IFN- γ -inducible genes (Sorace et al. (1995) *J. Leukocyte Biol.* 58:477-484). LRG-47 also has
20 homology to Tgtp and may be a GTP-binding protein (Sorace et al., *supra*).

RC10-II (Genbank Accession Number D21800) is gene that encodes the RC10-II subunit of the 20S proteasome of rat embryonic brain (Nishimura et al. (1993) *FEBS Lett.*
25 336:462-66). It has been suggested that RC10-II is a proteasomal subunit that is required for expression of tryptic activity (Nishimura et al., *supra*).

Stra13 (Genbank Accession Number AF010305) is a retinoic acid-inducible gene that encodes a basic helix-
30 loop-helix protein (Boudjelal et al. (1997) *Genes Dev.* 11:2052-65). Stra13 may act as a repressor of activated

transcription and is thought to play a role in neuronal differentiation (Boudjelal et al., *supra*).

Various aspects of the invention are described in further detail in the following subsections.

5

I. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode LIG46 or LIG56 proteins or biologically active portions thereof, as well as nucleic acid molecules which can be used as hybridization probes to identify LIG46 or LIG56-encoding nucleic acid molecules (e.g., human LIG46 or human LIG56) and fragments for use as PCR primers for the amplification or mutation of LIG46 or LIG56 nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

This section describes various LIG46 and LIG56 nucleic acid molecules. Of course, isolated nucleic acid molecules encoding all or part of Tgtp, LRG-47, RC10-II, and Stra13 are useful in the methods of the invention, e.g., methods for identifying compounds which modulate a body weight disorder. Thus, a nucleic acid molecule encompassing a sequence encoding all or part of Tgtp, LRG-47, RC10-II, or Stra13 (or a nucleic acid molecule encompassing all or part of the regulatory region of a Tgtp, LRG-47, RC10-II, or Stra13 gene) can be used to create recombinant cells that can be used in screening assays. In addition, nucleic acid molecules encoding Tgtp, LRG-47, RC10-II, and Stra13 can be

used to create transgenic mice which overexpress one or more of Tgtp, LRG-47, RC10-II, and Stra13. Such transgenic mice are useful in elucidating the role of these genes in body weight regulation. Thus, the methods described in this

5 section can be used to prepare and manipulate Tgtp, LRG-47, RC10-II, and Stra13 nucleic acid molecules as well as human homologues of Tgtp, LRG-47, RC10-II, and Stra13.

An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are

10 present in the natural source of the nucleic acid.

Preferably, an "isolated" nucleic acid is free of sequences (preferably protein encoding sequences) which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the

15 organism from which the nucleic acid is derived. For example, in various embodiments, the isolated LIG46 or LIG56 nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide

sequences which naturally flank the nucleic acid molecule in

20 genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical

25 precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5 or SEQ ID NO:7, or a complement of any of these nucleotide sequences, can be

30 isolated using standard molecular biology techniques and the sequence information provided herein. Using all or portion

of the nucleic acid sequences of SEQ ID NO:1, SEQ ID NO:3,
or all or a portion of the nucleic acid sequence of SEQ ID
NO:5 or SEQ ID NO:7, as a hybridization probe, LIG46 and
LIG56 nucleic acid molecules can be isolated using standard
5 hybridization and cloning techniques (e.g., as described in
Sambrook et al., eds., *Molecular Cloning: A Laboratory
Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring
Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

A nucleic acid of the invention can be amplified
10 using cDNA, mRNA or genomic DNA as a template and
appropriate oligonucleotide primers according to standard
PCR amplification techniques. The nucleic acid so amplified
can be cloned into an appropriate vector and characterized
by DNA sequence analysis. Furthermore, oligonucleotides
15 corresponding to LIG46 or LIG56 nucleotide sequences can be
prepared by standard synthetic techniques, e.g., using an
automated DNA synthesizer.

The isolated nucleic acid molecules of the invention
comprise a nucleic acid molecule which is a complement of
20 the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3,
SEQ ID NO:5 or SEQ ID NO:7, or a portion thereof. A nucleic
acid molecule which is complementary to a given nucleotide
sequence is one which is sufficiently complementary to the
given nucleotide sequence that it can hybridize to the given
25 nucleotide sequence thereby forming a stable duplex.

Moreover, the nucleic acid molecule of the invention
can comprise only a portion of a nucleic acid sequence
encoding LIG46 or LIG56, for example, a fragment which can
be used as a probe or primer or a fragment encoding a
30 biologically active portion of LIG46 or LIG56. The
nucleotide sequence determined from the cloning of the

murine LIG46 gene and the murine LIG56 gene allows for the generation of probes and primers designed for use in identifying and/or cloning LIG46 or LIG56 homologues in other cell types, e.g., from other tissues, as well as LIG46 and LIG56 homologues from other mammals, e.g., humans. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 50, 75, 100, 125, 150, 175, 200, 250, 300, 350 or 400 consecutive nucleotides of the sense or anti-sense sequence of SEQ ID NO:1 or SEQ ID NO:3, or of a naturally occurring mutant of SEQ ID NO:1 or SEQ ID NO:3, or sense or anti-sense sequence of SEQ ID NO:5 or SEQ ID NO:7, or of a naturally occurring mutant of SEQ ID NO:5 or SEQ ID NO:7.

Probes based on the LIG46 or LIG56 nucleotide sequence can be used to detect transcripts or genomic sequences encoding the same or related proteins (e.g., human homologues). The probe comprises a label group attached thereto, e.g., a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which mis-express a LIG46 or LIG56 protein, such as by measuring a level of a LIG46 or LIG56-encoding nucleic acid in a sample of cells from a subject, e.g., detecting LIG46 or LIG56 mRNA levels or determining whether a genomic LIG46 or LIG56 gene has been mutated or deleted.

A nucleic acid fragment encoding a "biologically active portion of LIG46" can be prepared by isolating a portion of SEQ ID NO:1 or SEQ ID NO:3 which encodes a

polypeptide having a LIG46 biological activity, expressing the encoded portion of LIG46 (e.g., by recombinant expression *in vitro*) and assessing the activity of the encoded portion of LIG46. For example, a nucleic acid
5 fragment encoding a biologically active portion of LIG46 includes a galactosyltransferase-like domain, e.g., SEQ ID NO:___.

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence of SEQ ID
10 NO:1 or SEQ ID NO:3 due to degeneracy of the genetic code and thus encode the same LIG46 protein as that encoded by the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:3.

In addition to the LIG46 nucleotide sequences shown in SEQ ID NO:1 and SEQ ID NO:3, it will be appreciated by
15 those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of LIG46 may exist within a population. Such genetic polymorphism in the LIG46 gene may exist among individuals within a population due to natural allelic variation. As used
20 herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a LIG46 protein, preferably a mammalian LIG46 protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the
25 LIG46 gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in LIG46 that are the result of natural allelic variation and that do not alter the functional activity of LIG46 are intended to be within the scope of the invention.

30 A nucleic acid fragment encoding a "biologically active portion of LIG56" can be prepared by isolating a

portion of SEQ ID NO:5 or SEQ ID NO:7 which encodes a polypeptide having a LIG56 biological activity, expressing the encoded portion of LIG56 protein (e.g., by recombinant expression *in vitro*) and assessing the activity of the
5 encoded portion of LIG56. For example, a nucleic acid fragment encoding a biologically active portion of LIG56 includes a GTP binding protein-like domain, e.g., SEQ ID NO:___.

The invention further encompasses nucleic acid
10 molecules that differ from the nucleotide sequence of SEQ ID NO:5 or SEQ ID NO:7 due to degeneracy of the genetic code and thus encode the same LIG56 protein as that encoded by the nucleotide sequence shown in SEQ ID NO:5 or SEQ ID NO:7.

In addition to the murine LIG56 nucleotide sequence
15 shown in SEQ ID NO:5 and SEQ ID NO:7, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of LIG56 may exist within a population. Such genetic polymorphism in the LIG56 gene may exist among individuals within a
20 population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a LIG56 protein, preferably a mammalian LIG56 protein. Such natural allelic variations can typically
25 result in 1-5% variance in the nucleotide sequence of the LIG56 gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in LIG56 that are the result of natural allelic variation and that do not alter the functional activity of LIG56 are intended to be within
30 the scope of the invention.

Moreover, nucleic acid molecules encoding LIG46 or

LIG56 proteins from other species (LIG46 or LIG56 homologues), which have a nucleotide sequence which differs from that of the murine gene, are intended to be within the scope of the invention. Nucleic acid molecules

5 corresponding to natural allelic variants and homologues of the LIG46 or LIG56 cDNA of the invention can be isolated based on their identity to the LIG46 or LIG56 nucleic acids disclosed herein using the murine cDNAs, or a portion thereof, as a hybridization probe according to standard

10 hybridization techniques under stringent hybridization conditions. For example, a soluble LIG46 cDNA can be isolated based on its identity to murine membrane-bound LIG46. Likewise, a membrane-bound human LIG56 cDNA can be isolated based on its identity to soluble LIG56.

15 Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 300 (325, 350, 375, 400, 425, 450, 500, 550, 600, 650, 700, 800, 900, 1000, or 1200) nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising
20 the nucleotide sequence, preferably the coding sequence, of SEQ ID NO:1, or SEQ ID NO:3, or SEQ ID NO:5 or SEQ ID NO:7.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences
25 at least 60% (65%, 70%, preferably 75%) identical to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred,
30 non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC)

at about 45°C, followed by one or more washes in 0.2 X SSC,
0.1% SDS at 50-65°C. Preferably, an isolated nucleic acid
molecule of the invention that hybridizes under stringent
conditions to the sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ
5 ID NO:5 or SEQ ID NO:7 corresponds to a naturally-occurring
nucleic acid molecule. As used herein, a "naturally-
occurring" nucleic acid molecule refers to an RNA or DNA
molecule having a nucleotide sequence that occurs in nature
(e.g., encodes a natural protein).

10 In addition to naturally-occurring allelic variants
of the LIG46 or LIG56 sequence that may exist in the
population, the skilled artisan will further appreciate that
changes can be introduced by mutation into the nucleotide
sequences disclosed herein, thereby leading to changes in
15 the amino acid sequence of the encoded LIG46 or LIG56
protein, without altering the functional ability of the
LIG46 or LIG56 protein. For example, one can make
nucleotide substitutions leading to amino acid substitutions
at "non-essential" amino acid residues. A "non-essential"
20 amino acid residue is a residue that can be altered from the
wild-type sequence of LIG46 or LIG56 without altering the
biological activity, whereas an "essential" amino acid
residue is required for biological activity. For example,
amino acid residues that are conserved among the LIG46 or
25 LIG56 proteins of various species are predicted to be
particularly unamenable to alteration.

For example, preferred LIG46 proteins of the present
invention retain amino acids that are conserved among
galactosyltransferases. Such conserved domains are less
30 likely to be amenable to mutation. Other amino acid
residues, however, (e.g., those that are not conserved or

only semi-conserved among LIG46 or LIG56 of various species) may not be essential for activity and thus are likely to be amenable to alteration.

Accordingly, another aspect of the invention
5 pertains to nucleic acid molecules encoding LIG46 or LIG56 proteins that contain changes in amino acid residues that are not essential for activity. Such LIG46 or LIG56 proteins differ in amino acid sequence from those disclosed herein yet retain biological activity. In one embodiment,
10 the isolated nucleic acid molecule includes a nucleotide sequence encoding a protein that includes an amino acid sequence that is at least about 45% identical, 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:6.

15 An isolated nucleic acid molecule encoding a LIG46 or LIG56 protein having a sequence which differs from that disclosed herein can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence disclosed herein such that one or more
20 amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or
25 more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the
30 art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side

chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline,

5 phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in LIG46 or LIG56 is preferably replaced with
10 another amino acid residue from the same side chain family.

Alternatively, mutations can be introduced randomly along all or part of a LIG46 or LIG56 coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for LIG46 or LIG56 biological activity to identify
15 mutants that retain activity. Following mutagenesis, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

The present invention encompasses antisense nucleic acid molecules, i.e., molecules which are complementary to a
20 sense nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an
25 entire LIG46 or LIG56 coding strand, or to only a portion thereof, e.g., all or part of the protein coding region (or open reading frame). An antisense nucleic acid molecule can be antisense to a noncoding region of the coding strand of a nucleotide sequence encoding LIG46 or LIG56. The noncoding
30 regions ("5' and 3' untranslated regions") are the 5' and 3' sequences which flank the coding region and are not

translated into amino acids.

Given the coding strand sequences encoding LIG46 or LIG56 disclosed herein (e.g., SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, and SEQ ID NO:7), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of LIG46 or LIG56 mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of LIG46 or LIG56 mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of LIG46 or LIG56 mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-

isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-

5 methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine.

Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding the protein of interest to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention include direct injection at a tissue site.

Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they

5 specifically bind to receptors or antigens expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using
10 the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

15 An antisense nucleic acid molecule of the invention can be an α -anomeric nucleic acid molecule. An α -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -units, the strands run parallel to each other (Gaultier et
20 al. (1987) *Nucleic Acids. Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-O-methylribonucleotide (Inoue et al. (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al. (1987) *FEBS Lett.* 215:327-330).

25 The invention also encompasses ribozymes. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes (described in
30 Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave LIG46 or LIG56 mRNA transcripts

to thereby inhibit translation of LIG46 or LIG56 mRNA. A ribozyme having specificity for a LIG46 or LIG56-encoding nucleic acid can be designed based upon the nucleotide sequence of a LIG46 or LIG56 cDNA disclosed herein. For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a LIG46 or LIG56-encoding mRNA. See, e.g., Cech et al. U.S. Patent No. 4,987,071; and Cech et al. U.S. Patent No. 5,116,742. Alternatively, LIG46 or LIG56 mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel and Szostak (1993) *Science* 261:1411-1418.

The invention also encompasses nucleic acid molecules which form triple helical structures. For example, LIG46 or LIG56 gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the LIG46 or LIG56 (e.g., the LIG46 or LIG56 promoter and/or enhancers) to form triple helical structures that prevent transcription of the LIG46 or LIG56 gene in target cells. See generally, Helene (1991) *Anticancer Drug Des.* 6(6):569-84; Helene (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14(12):807-15.

In preferred embodiments, the nucleic acid molecules of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup et al. (1996) *Bioorganic & Medicinal*

Chemistry 4(1): 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, e.g., DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four
5 natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in
10 Hyrup et al. (1996) *supra*; Perry-O'Keefe et al. (1996) *Proc. Natl. Acad. Sci. USA* 93: 14670-675.

PNAs of LIG46 or LIG56 can be used therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific
15 modulation of gene expression by, e.g., inducing transcription or translation arrest or inhibiting replication. PNAs of LIG46 or LIG56 can also be used, e.g., in the analysis of single base pair mutations in a gene by, e.g., PNA directed PCR clamping; as 'artificial restriction
20 enzymes when used in combination with other enzymes, e.g., S1 nucleases (Hyrup (1996) *supra*; or as probes or primers for DNA sequence and hybridization (Hyrup (1996) *supra*; Perry-O'Keefe et al. (1996) *Proc. Natl. Acad. Sci. USA* 93: 14670-675).

25 In another embodiment, PNAs of LIG46 or LIG56 can be modified, e.g., to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the
30 art. For example, PNA-DNA chimeras of LIG46 or LIG56 can be generated which may combine the advantageous properties of

PNA and DNA. Such chimeras allow DNA recognition enzymes, e.g., RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996) *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996) *supra* and Finn et al. (1996) *Nucleic Acids Research* 24(17):3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs, e.g., 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used as a between the PNA and the 5' end of DNA (Mag et al. (1989) *Nucleic Acid Res.* 17:5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn et al. (1996) *Nucleic Acids Research* 24(17):3357-63). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser et al. (1975) *Bioorganic Med. Chem. Lett.* 5:1119-11124).

In other embodiments, the oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, e.g., Letsinger et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. W0 88/09810) or the blood-brain barrier (see, e.g., PCT Publication No. W0 89/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (see, e.g., Krol et

al. (1988) *Bio/Techniques* 6:958-976) or intercalating agents (see, e.g., Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, e.g., a peptide, hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

II. Isolated LIG46 Proteins, Isolated LIG56 Proteins, Anti-LIG46 antibodies, and Anti-LIG56 Antibodies

One aspect of the invention pertains to isolated LIG46 or LIG56 proteins, and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise anti-LIG46 or LIG56 antibodies. In one embodiment, native LIG46 or LIG56 proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, LIG46 or LIG56 proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a LIG46 or LIG56 protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

This section focusses on LIG46 and LIG56 polypeptides, antibodies, and their use. However, Tgtp, LRG-47, RC10-II, and Stra13 polypeptides and antibodies (and fragments or variants thereof) are useful in the methods of the invention as are fusion proteins which include all or a portion of Tgtp, LRG-47, RC10-II, or Stra13. Thus, the methods described in this section for the production and use of LIG46 and LIG56 polypeptides and variants thereof apply to Tgtp, LRG-47, RC10-II, and Stra13. Antibodies directed against Tgtp, LRG-47, RC10-II, or Stra13 are useful in the

method of the invention. For example, such antibodies can be used to measure expression of Tgtp, LRG-47, RC10-II, or Stral3 in screening assays designed to identify agents which modulate expression or activity of Tgtp, LRG-47, RC10-II, or Stral3. The description methods for preparing and characterizing anti-LIG46 and anti-LIG56 antibodies presented below can be applied to antibodies directed against Tgtp, LRG-47, RC10-II, or Stral3.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein of interest is derived (e.g., LIG46 or LIG56), or substantially free from chemical precursors or other chemicals when chemically synthesized.

The language "substantially free of cellular material" includes preparations in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. Thus, LIG46 or LIG56 protein that is substantially free of cellular material includes preparations of LIG46 or LIG56 protein having less than about 30%, 20%, 10%, or 5% (by dry weight) of non-LIG46 or LIG56 protein (also referred to herein as a "contaminating protein"). When the LIG46 or LIG56 protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, 10%, or 5% of the volume of the protein preparation. When LIG46 or LIG56 protein is produced by chemical synthesis, it is preferably substantially free of chemical precursors or other chemicals, i.e., it is separated from chemical precursors or other chemicals which are involved in the synthesis of the

protein. Accordingly such preparations of LIG46 or LIG56 protein have less than about 30%, 20%, 10%, 5% (by dry weight) of chemical precursors or non-LIG46 or LIG56 chemicals.

5 Biologically active portions of a LIG46 or LIG56 protein include peptides comprising amino acid sequences sufficiently identical to or derived from the amino acid sequence of the LIG46 or LIG56 protein, which include less amino acids than the full length LIG46 or LIG56 proteins,
10 and exhibit at least one activity of a LIG46 or LIG56 protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the LIG46 or LIG56 protein. A biologically active portion of a LIG46 or LIG56 protein can be a polypeptide which is, for example,
15 10, 25, 50, 100 or more amino acids in length. Preferred biologically active polypeptides include one or more identified LIG46 or LIG56 structural domains.

Moreover, other biologically active portions, in which other regions of the protein are deleted, can be
20 prepared by recombinant techniques and evaluated for one or more of the functional activities of a native LIG46 or LIG56 protein.

Preferred LIG46 and LIG56 proteins have or are substantially identical to the amino acid sequences
25 disclosed herein. Preferred proteins are substantially identical to those disclosed herein and retain the functional activity of the protein yet differ in amino acid sequence due to natural allelic variation or mutagenesis.

Accordingly, a useful LIG46 protein is a protein
30 which includes an amino acid sequence at least about 45%, preferably 55%, 65%, 75%, 85%, 95%, or 99% identical to the

amino acid sequence of SEQ ID NO:2 (or SEQ ID NO:4) and retains the functional activity of the LIG46 protein of SEQ ID NO:2 (or SEQ ID NO:4). In other instances, the LIG46 protein is a protein having an amino acid sequence 55%, 65%, 5 75%, 85%, 95%, or 98% identical to a portion of LIG46 having homology to a galactosyltransferase (e.g., amino acids 192-353, 142-184, 201-296, 289-347, 140-183, 367-391, 177-266, 299-343, or 140-184 of SEQ ID NO:2) or a neurogenic secreted signalling protein (e.g., amino acids 200-291, 270-354, 144-10 183, 380-394, or 211-248 of SEQ ID NO:2). In a preferred embodiment, the LIG46 protein retains a functional activity of the LIG46 protein of SEQ ID NO:2 (or SEQ ID NO:4).

A useful LIG56 protein is a protein which includes an amino acid sequence at least about 45%, preferably 55%, 15 65%, 75%, 85%, 95%, or 99% identical to the amino acid sequence of SEQ ID NO:6 and retains the functional activity of the LIG46 protein of SEQ ID NO:6.

To determine the percent identity of two amino acid sequences or of two nucleic acids, the sequences are aligned 20 for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or 25 nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is 30 a function of the number of identical positions shared by the sequences (i.e., % identity = # of identical

positions/total # of positions x 100).

The determination of percent homology between two sequences can be accomplished using a mathematical algorithm. A preferred, non-limiting example of a

5 mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin and Altschul (1990) *Proc. Nat'l Acad. Sci. USA* 87:2264-2268, modified as in Karlin and Altschul (1993) *Proc. Nat'l Acad. Sci. USA* 90:5873-5877. Such an algorithm is incorporated into the

10 NBLAST and XBLAST programs of Altschul, et al. (1990) *J. Mol. Biol.* 215:403-410. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to LIG46 or LIG56 nucleic acid molecules of the invention. BLAST

15 protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to LIG46 or LIG56 protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in

20 Altschul et al., (1997) *Nucleic Acids Res.* 25:3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See <http://www.ncbi.nlm.nih.gov>. Another preferred, non-limiting example of a mathematical

25 algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, CABIOS (1989). Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing

30 amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used.

The percent identity between two sequences can be determined using techniques similar to those described above, with or without allowing gaps. In calculating percent identity, only exact matches are counted.

5 The invention also provides LIG46 or LIG56 chimeric or fusion proteins. As used herein, a LIG46 or LIG56 "chimeric protein" or "fusion protein" comprises a LIG46 or LIG56 polypeptide operatively linked to a non-LIG46 or LIG56 polypeptide. A "LIG46 or LIG56 polypeptide" refers to a
10 polypeptide having an amino acid sequence corresponding to LIG46 or LIG56, whereas a "non-LIG46 or LIG56 polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially identical to the LIG46 or LIG56 protein, e.g., a protein
15 which is different from the LIG46 or LIG56 protein and which is derived from the same or a different organism. Within a LIG46 or LIG56 fusion protein the LIG46 or LIG56 polypeptide can correspond to all or a portion of a LIG46 or LIG56 protein, preferably at least one biologically active portion
20 of a LIG46 or LIG56 protein. Within the fusion protein, the term "operatively linked" is intended to indicate that the LIG46 or LIG56 polypeptide and the non-LIG46 or LIG56 polypeptide are fused in-frame to each other. The non-LIG46 or LIG56 polypeptide can be fused to the N-terminus or C-
25 terminus of the LIG46 or LIG56 polypeptide.

One useful fusion protein is a GST-LIG46 or LIG56 fusion protein in which the LIG46 or LIG56 sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant
30 LIG46 or LIG56.

In another embodiment, the fusion protein is a LIG46

protein containing a heterologous signal sequence at its N-terminus. For example, the native LIG46 signal sequence (i.e., about amino acids 1 to 32 of SEQ ID NO:2) can be removed and replaced with a signal sequence from another protein. In certain host cells (e.g., mammalian host cells), expression and/or secretion of LIG46 can be increased through use of a heterologous signal sequence. For example, the gp67 secretory sequence of the baculovirus envelope protein can be used as a heterologous signal sequence (*Current Protocols in Molecular Biology*, Ausubel et al., eds., John Wiley & Sons, 1992). Other examples of eukaryotic heterologous signal sequences include the secretory sequences of melittin and human placental alkaline phosphatase (Stratagene; La Jolla, California). In yet another example, useful prokaryotic heterologous signal sequences include the phoA secretory signal (*Molecular cloning*, Sambrook et al, second edition, Cold spring harbor laboratory press, 1989) and the protein A secretory signal (Pharmacia Biotech; Piscataway, New Jersey).

In yet another embodiment, the fusion protein is an LIG46 or LIG56-immunoglobulin fusion protein in which all or part of LIG46 or LIG56 is fused to sequences derived from a member of the immunoglobulin protein family. The LIG46-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a LIG56 ligand and a LIG56 protein on the surface of a cell, to thereby suppress LIG56-mediated signal transduction *in vivo*. The LIG56-immunoglobulin fusion proteins can be used to affect the bioavailability of a LIG56 cognate ligand. Moreover, the LIG56-immunoglobulin fusion proteins of the

invention can be used as immunogens to produce LIG56 antibodies in a subject, to purify LIG56 ligands and in screening assays to identify molecules which inhibit the interaction of LIG56 with a LIG56 ligand. LIG46 fusion
5 proteins can be used in an analogous manner.

Preferably, a LIG46 or LIG56 chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-
10 frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid
15 undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give
20 rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, e.g., *Current Protocols in Molecular Biology*, Ausubel et al. eds., John Wiley & Sons: 1992). Moreover, many expression vectors are
25 commercially available that already encode a fusion moiety (e.g., a GST polypeptide). An LIG46- or LIG56-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the LIG46 or LIG56 protein.

30 The present invention also pertains to variants of the LIG46 or LIG56 proteins which function as either LIG46

or LIG56 agonists (mimetics) or as LIG46 or LIG56 antagonists. Variants of the LIG46 or LIG56 protein can be generated by mutagenesis, e.g., discrete point mutation or truncation of the LIG46 or LIG56 protein. An agonist of the LIG46 or LIG56 protein can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of the LIG46 or LIG56 protein. An antagonist of the LIG46 or LIG56 protein can inhibit one or more of the activities of the naturally occurring form of the LIG46 or LIG56 protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the LIG46 or LIG56 protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. Treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein can have fewer side effects in a subject relative to treatment with the naturally occurring form of the LIG46 or LIG56 proteins.

Variants of the LIG46 or LIG56 protein which function as either LIG46 or LIG56 agonists (mimetics) or as LIG46 or LIG56 antagonists can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of the LIG46 or LIG56 protein for LIG46 or LIG56 protein agonist or antagonist activity. In one embodiment, a variegated library of LIG46 or LIG56 variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of LIG46 or LIG56 variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential LIG46 or LIG56 sequences is expressible as

individual polypeptides, or alternatively, as a set of
larger fusion proteins (e.g., for phage display) containing
the set of LIG46 or LIG56 sequences therein. There are a
variety of methods which can be used to produce libraries of
5 potential LIG46 or LIG56 variants from a degenerate
oligonucleotide sequence. Chemical synthesis of a
degenerate gene sequence can be performed in an automatic
DNA synthesizer, and the synthetic gene then ligated into an
appropriate expression vector. Use of a degenerate set of
10 genes allows for the provision, in one mixture, of all of
the sequences encoding the desired set of potential LIG46 or
LIG56 sequences. Methods for synthesizing degenerate
oligonucleotides are known in the art (see, e.g., Narang
(1983) *Tetrahedron* 39:3; Itakura et al. (1984) *Annu. Rev.*
15 *Biochem.* 53:323; Itakura et al. (1984) *Science* 198:1056; Ike
et al. (1983) *Nucleic Acid Res.* 11:477).

In addition, libraries of fragments of the LIG46 or
LIG56 protein coding sequence can be used to generate a
variegated population of LIG46 or LIG56 fragments for
20 screening and subsequent selection of variants of a LIG46 or
LIG56 protein. In one embodiment, a library of coding
sequence fragments can be generated by treating a double
stranded PCR fragment of a LIG46 or LIG56 coding sequence
with a nuclease under conditions wherein nicking occurs only
25 about once per molecule, denaturing the double stranded DNA,
renaturing the DNA to form double stranded DNA which can
include sense/antisense pairs from different nicked
products, removing single stranded portions from reformed
duplexes by treatment with S1 nuclease, and ligating the
30 resulting fragment library into an expression vector. By
this method, an expression library can be derived which

encodes N-terminal and internal fragments of various sizes of the LIG46 or LIG56 protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of LIG46 or LIG56 proteins. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify LIG46 or LIG56 variants (Arkin and Yourvan (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave et al. (1993) *Protein Engineering* 6(3):327-331).

An isolated LIG46 or LIG56 protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind LIG46 or LIG56 using standard techniques for polyclonal and monoclonal antibody preparation. The full-length LIG46 or LIG56 protein can be used or, alternatively, the invention provides antigenic peptide fragments of LIG46 or LIG56 for use as immunogens. The antigenic peptide of LIG46 or LIG56 comprises at least 8

(preferably 10, 15, 20, or 30) amino acid residues of the amino acid sequence shown in SEQ ID NO:2 and encompasses an epitope of LIG46 or LIG56 such that an antibody raised against the peptide forms a specific immune complex with
5 LIG46 or LIG56.

Preferred epitopes encompassed by the antigenic peptide are regions of LIG46 or LIG56 that are located on the surface of the protein, e.g., hydrophilic regions. Hydrophilic regions and antigenic regions can be identified
10 using standard analytical tools well-known to those skilled in the art.

A LIG46 or LIG56 immunogen typically is used to prepare antibodies by immunizing a suitable subject, (e.g., rabbit, goat, mouse or other mammal) with the immunogen. An
15 appropriate immunogenic preparation can contain, for example, recombinantly expressed LIG46 or LIG56 protein or a chemically synthesized LIG46 or LIG56 polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar
20 immunostimulatory agent. Immunization of a suitable subject with an immunogenic LIG46 or LIG56 preparation induces a polyclonal anti-LIG46 or LIG56 antibody response.

Accordingly, another aspect of the invention pertains to anti-LIG46 or LIG56 antibodies. The term
25 "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antigen binding site which specifically binds an antigen, such as LIG46 or LIG56. A molecule which specifically binds to LIG46 or
30 LIG56 is a molecule which binds LIG46 or LIG56, but does not substantially bind other molecules in a sample, e.g., a

biological sample, which naturally contains LIG46 or LIG56.

Examples of immunologically active portions of immunoglobulin molecules include F(ab) and F(ab')₂ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies that bind LIG46 or LIG56. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of LIG46 or LIG56. A monoclonal antibody composition thus typically displays a single binding affinity for a particular LIG46 or LIG56 protein with which it immunoreacts.

Polyclonal anti-LIG46 or LIG56 antibodies can be prepared as described above by immunizing a suitable subject with a LIG46 or LIG56 immunogen. The anti-LIG46 or LIG56 antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized LIG46 or LIG56. If desired, the antibody molecules directed against LIG46 or LIG56 can be isolated from the mammal (e.g., from the blood) and further purified by well-known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, e.g., when the anti-LIG46 or LIG56 antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975) *Nature* 256:495-497, the human B cell hybridoma technique (Kozbor et

al. (1983) *Immunol Today* 4:72), the EBV-hybridoma technique (Cole et al. (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing various antibodies
5 monoclonal antibody hybridomas is well known (see generally Current Protocols in Immunology (1994) Coligan et al. (eds.) John Wiley & Sons, Inc., New York, NY). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized
10 with a LIG46 or LIG56 immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds LIG46 or LIG56.

Any of the many well known protocols used for fusing
15 lymphocytes and immortalized cell lines can be applied for the purpose of generating an anti-LIG46 or LIG56 monoclonal antibody (see, e.g., Current Protocols in Immunology, *supra*; Galfre et al. (1977) *Nature* 266:55052; R.H. Kenneth, in *Monoclonal Antibodies: A New Dimension In Biological*
20 *Analyses*, Plenum Publishing Corp., New York, New York (1980); and Lerner (1981) *Yale J. Biol. Med.*, 54:387-402. Moreover, the ordinarily skilled worker will appreciate that there are many variations of such methods which also would be useful. Typically, the immortal cell line (e.g., a
25 myeloma cell line) is derived from the same mammalian species as the lymphocytes. For example, murine hybridomas can be made by fusing lymphocytes from a mouse immunized with an immunogenic preparation of the present invention with an immortalized mouse cell line, e.g., a myeloma cell
30 line that is sensitive to culture medium containing hypoxanthine, aminopterin and thymidine ("HAT medium"). Any

of a number of myeloma cell lines can be used as a fusion partner according to standard techniques, e.g., the P3-NS1/1-Ag4-1, P3-x63-Ag8.653 or Sp2/O-Ag14 myeloma lines. These myeloma lines are available from ATCC. Typically,
5 HAT-sensitive mouse myeloma cells are fused to mouse splenocytes using polyethylene glycol ("PEG"). Hybridoma cells resulting from the fusion are then selected using HAT medium, which kills unfused and unproductively fused myeloma cells (unfused splenocytes die after several days because
10 they are not transformed). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind LIG46 or LIG56, e.g., using a standard ELISA assay.

15 Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal anti-LIG46 or LIG56 antibody can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (e.g., an antibody phage display library) with LIG46 or LIG56 to
20 thereby isolate immunoglobulin library members that bind LIG46 or LIG56. Kits for generating and screening phage display libraries are commercially available (e.g., the Pharmacia Recombinant Phage Antibody System, Catalog No. 27-9400-01; and the Stratagene SurfZAP™ Phage Display Kit,
25 Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, U.S. Patent No. 5,223,409; PCT Publication No. WO 92/18619; PCT Publication No. WO 91/17271; PCT Publication
30 No. WO 92/20791; PCT Publication No. WO 92/15679; PCT Publication No. WO 93/01288; PCT Publication No. WO

92/01047; PCT Publication No. WO 92/09690; PCT Publication
No. WO 90/02809; Fuchs et al. (1991) *Bio/Technology* 9:1370-
1372; Hay et al. (1992) *Hum. Antibod. Hybridomas* 3:81-85;
Huse et al. (1989) *Science* 246:1275-1281; Griffiths et al.
5 (1993) *EMBO J* 12:725-734.

Additionally, recombinant anti-LIG46 or LIG56
antibodies, such as chimeric and humanized monoclonal
antibodies, comprising both human and non-human portions,
which can be made using standard recombinant DNA techniques,
10 are within the scope of the invention. Such chimeric and
humanized monoclonal antibodies can be produced by
recombinant DNA techniques known in the art, for example
using methods described in PCT Publication No. WO 87/02671;
European Patent Application 184,187; European Patent
15 Application 171,496; European Patent Application 173,494;
PCT Publication No. WO 86/01533; U.S. Patent No. 4,816,567;
European Patent Application 125,023; Better et al. (1988)
Science 240:1041-1043; Liu et al. (1987) *Proc. Natl. Acad.*
Sci. USA 84:3439-3443; Liu et al. (1987) *J. Immunol.*
20 139:3521-3526; Sun et al. (1987) *Proc. Natl. Acad. Sci. USA*
84:214-218; Nishimura et al. (1987) *Canc. Res.* 47:999-1005;
Wood et al. (1985) *Nature* 314:446-449; and Shaw et al.
(1988) *J. Natl. Cancer Inst.* 80:1553-1559; Morrison, (1985)
Science 229:1202-1207; Oi et al. (1986) *Bio/Techniques*
25 4:214; U.S. Patent 5,225,539; Jones et al. (1986) *Nature*
321:552-525; Verhoeyan et al. (1988) *Science* 239:1534; and
Beidler et al. (1988) *J. Immunol.* 141:4053-4060.

An anti-LIG46 or LIG56 antibody (e.g., monoclonal
antibody) can be used to isolate LIG46 or LIG56 by standard
30 techniques, such as affinity chromatography or
immunoprecipitation. An anti-LIG46 or LIG56 antibody can

facilitate the purification of natural LIG46 or LIG56 from cells and of recombinantly produced LIG46 or LIG56 expressed in host cells. Moreover, an anti-LIG46 or LIG56 antibody can be used to detect LIG46 or LIG56 protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the LIG46 or LIG56 protein. Anti-LIG46 or LIG56 antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

Completely human antibodies are particularly desirable for therapeutic treatment of human patients. Such antibodies can be produced using transgenic mice which are incapable of expressing endogenous immunoglobulin heavy and light chains genes, but which can express human heavy and

light chain genes. The transgenic mice are immunized in the normal fashion with a selected antigen. Monoclonal antibodies directed against the antigen can be obtain using conventional hybridoma technology. The human immunoglobulin transgenes of harbored by the transgenic mice rearrange during B cell differentiation, subsequently undergo class switching and somatic mutation. Thus, using such a technique, it is possible to produce therapeutically useful IgG, IgA and IgE antibodies. For an overview of this technology for producing human antibodies, see Lonberg and Huszar (1995, *Int. Rev. Immunol.* 13:65-93). For a detailed discussion of this technology for producing human antibodies and human monoclonal antibodies and protocols for producing such antibodies, see, e.g., U.S. Patent 5,625,126; U.S. Patent 5,633,425; U.S. Patent 5,569,825; U.S. Patent 5,661,016; and U.S. Patent 5,545,806. Human antibodies directed against a selected antigen can be provided by Abgenix, Inc. (Fremont, CA) and GenPharm, Inc. (Palo Alto, CA).

III. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding LIG46 or LIG56 (or a portion thereof).

The techniques described below can also be applied to host cells and vectors used to express Tgtp, LRG-47, RC10-II, and Stra13 for use in the production of recombinant protein or transgenic animals. Thus, although the this section refers to LIG46 and LIG56, the methods described can be applied to Tgtp, LRG-47, RC10-II, and Stra13.

As used herein, the term "vector" refers to a

nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be
5 ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of
10 replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome.

Moreover, certain vectors, expression vectors, are capable
15 of directing the expression of genes to which they are operatively linked. In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids (vectors). However, the invention is intended to include such other forms of expression vectors, such as
20 viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable
25 for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a
30 recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked

to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term

5 "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San
10 Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). It will be
15 appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to
20 thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., LIG46 or LIG56 proteins, mutant forms of LIG46 or LIG56, fusion proteins, etc.).

The recombinant expression vectors of the invention
25 can be designed for expression of LIG46 or LIG56 in prokaryotic or eukaryotic cells, e.g., bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression*
30 *Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression

vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al., (1988) *Gene* 69:301-315) and pET 11d (Studier et al., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89). Target gene expression

from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gnl0-lac fusion promoter mediated by
5 a coexpressed viral RNA polymerase (T7 gnl). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident λ prophage harboring a T7 gnl gene under the transcriptional control of the lacUV 5 promoter.

10 One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San
15 Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada et al. (1992) *Nucleic Acids Res.* 20:2111-
20 2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the LIG46 or LIG56 expression vector is a yeast expression vector. Examples of vectors
25 for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari et al. (1987) *EMBO J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz et al. (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), and picZ (InVitrogen Corp, San Diego, CA).

30 Alternatively, LIG46 or LIG56 can be expressed in insect cells using baculovirus expression vectors.

Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al. (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-5 39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed (1987) *Nature* 329:840) and pMT2PC 10 (Kaufman et al. (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For 15 other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook et al. (*supra*).

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the 20 nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al. (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji et al. (1983) *Cell* 33:729-740; 30 Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and

Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund et al. (1985) *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α -fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

10 The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for
15 expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to LIG46 or LIG56 mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA
20 molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant
25 plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene
30 expression using antisense genes see Weintraub et al. (*Reviews - Trends in Genetics*, Vol. 1(1) 1986).

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, LIG46 or LIG56 protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (*supra*), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells

may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding LIG46 or LIG56 or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (i.e., express) LIG46 or LIG56 protein. Accordingly, the invention further provides methods for producing LIG46 or LIG56 protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding LIG46 or LIG56 has been introduced) in a suitable medium such that LIG46 or LIG56 protein is produced. In another embodiment, the method further comprises isolating LIG46 or LIG56 from the medium or the host cell.

The host cells of the invention can also be used to produce non-human transgenic animals which over-express a protein of interest. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which a nucleic acid molecule which directs high level expression of LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stral3 has been introduced. Such host cells can then

be used to create non-human transgenic animals in which
LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 sequences
have been introduced into their genome or homologous
recombinant animals in which endogenous LIG46, LIG56, Tgtp,
5 LRG-47, RC10-II, or Stra13 sequences have been altered.
Such animals are useful for studying the function and/or
activity of LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13
and for identifying and/or evaluating modulators of LIG46 or
LIG56 activity. As used herein, a "transgenic animal" is a
10 non-human animal, preferably a mammal, more preferably a
rodent such as a rat or mouse, in which one or more of the
cells of the animal includes a transgene. Other examples of
transgenic animals include non-human primates, sheep, dogs,
cows, goats, chickens, amphibians, etc. A transgene is
15 exogenous DNA which is integrated into the genome of a cell
from which a transgenic animal develops and which remains in
the genome of the mature animal, thereby directing the
expression of an encoded gene product in one or more cell
types or tissues of the transgenic animal. As used herein,
20 an "homologous recombinant animal" is a non-human animal,
preferably a mammal, more preferably a mouse, in which an
endogenous LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13
gene has been altered by homologous recombination between
the endogenous gene and an exogenous DNA molecule introduced
25 into a cell of the animal, e.g., an embryonic cell of the
animal, prior to development of the animal.

A transgenic animal of the invention can be created
by introducing a nucleic acid molecule encoding a desired
protein into the male pronuclei of a fertilized oocyte,
30 e.g., by microinjection, retroviral infection, and allowing
the oocyte to develop in a pseudopregnant female foster

animal. The cDNA sequence can be introduced as a transgene into the genome of a non-human animal. Alternatively, a human homologue of the LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 gene can be isolated based on hybridization to the murine LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 cDNA and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, U.S. Patent No. 4,873,191 and in Hogan, *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of the mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 can further be bred to other transgenic animals carrying other transgenes.

To create an homologous recombinant animal, a vector is prepared which contains at least a portion of a LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 gene into which a deletion, addition or substitution has been introduced to

thereby alter, e.g., functionally disrupt, the gene. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous gene is mutated or otherwise altered but still encodes functional protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous protein). In the homologous recombination vector, the altered portion of the gene is flanked at its 5' and 3' ends by additional nucleic acids of the gene to allow for homologous recombination to occur between the exogenous gene carried by the vector and an endogenous gene in an embryonic stem cell.

The additional flanking nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see, e.g., Thomas and Capecchi (1987) *Cell* 51:503 for a description of homologous recombination vectors). The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced gene has homologously recombined with the endogenous gene are selected (see, e.g., Li et al. (1992) *Cell* 69:915). The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (see, e.g., Bradley in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo

brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley (1991) *Current Opinion in Bio/Technology* 2:823-829 and in PCT Publication Nos. WO 90/11354, WO 91/01140, WO 92/0968, and WO 93/04169.

In another embodiment, transgenic non-humans animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, e.g., Lakso et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman et al. (1991) *Science* 251:1351-1355. If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut et al. (1997) *Nature* 385:810-813 and PCT Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can

be isolated and induced to exit the growth cycle and enter G₀ phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal.

The offspring borne of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

IV. Pharmaceutical Compositions

The LIG46 and LIG56 nucleic acid molecules, LIG46 and LIG56 proteins, and anti-LIG46 and anti- LIG56 antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions suitable for administration as can various modulators of LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 expression or activity.

Therapeutic compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active

compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration

5 include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following

10 components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as
15 ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or
20 sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble)
25 or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF; Parsippany, NJ) or phosphate buffered saline (PBS).
30 In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must

be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example,
5 water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size
10 in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to
15 include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example,
20 aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., a LIG46 or LIG56 protein or anti-LIG46 or LIG56 antibody) in the required amount in an appropriate solvent with one or a combination
25 of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case
30 of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum

drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent
5 or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also
10 be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills,
15 capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn
20 starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are
25 delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal
30 administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are

generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or

5 suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases
10 such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release
15 formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be
20 apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as
25 pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of
30 administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as

unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (U.S. Patent 5,328,470) or by stereotactic injection (see, e.g., Chen et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057).

The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g. retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

V. Uses and Methods of the Invention

The LIG46, LIG56, Tgtp, LRG-47, RC10-II, and Stra13 nucleic acid molecules, proteins, protein homologues, and

antibodies described herein can be used in one or more of the following methods: a) screening assays; b) detection assays (e.g., chromosomal mapping, tissue typing, forensic biology); and c) methods of treatment (e.g., therapeutic and prophylactic). The isolated nucleic acid molecules of the invention can be used to express LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 protein (e.g., via a recombinant expression vector in a host cell in gene therapy applications or transgenic animals), to detect LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 mRNA (e.g., in a biological sample) or a genetic lesion in a LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 gene, and to modulate LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 activity or expression. In addition, LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 proteins can be used to screen drugs or compounds which modulate LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 activity or expression as well as to treat disorders characterized by insufficient or excessive production of LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 protein or production of LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 protein forms which have an undesirable level of activity compared to the wild type protein. In addition, the anti-LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 antibodies of the invention can be used to detect and isolate LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 proteins and modulate LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 activity.

This invention further pertains to novel agents identified by the above-described screening assays and uses thereof for treatments as described herein.

A. Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, i.e., candidate or test compounds or agents (e.g., peptides, peptidomimetics, small molecules or other drugs) which bind to a LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 protein and/or have a stimulatory or inhibitory effect on, for example, LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 expression or activity.

The invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of a LIG46 protein or polypeptide or biologically active portion thereof. Other embodiments entail the use of a soluble form of LIG46.

The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam (1997) *Anticancer Drug Des.* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt et al. (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909; Erb et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann et al. (1994). *J. Med. Chem.* 37:2678; Cho et al. (1993)

Science 261:1303; Carrell et al. (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell et al. (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and Gallop et al. (1994) *J. Med. Chem.* 37:1233.

5 Libraries of compounds may be presented in solution (e.g., Houghten (1992) *Bio/Techniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (U.S. Patent No. 5,223,409), spores (Patent Nos. 5,571,698; 5,403,484; and 5,223,409),
10 plasmids (Cull et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390; Devlin (1990) *Science* 249:404-406; Cwirla et al. (1990) *Proc. Natl. Acad. Sci.* 87:6378-6382; and Felici (1991) *J. Mol. Biol.* 222:301-310).

15 The invention includes assays employing soluble LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13. Such assays entail contacting a LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 protein or biologically active portion thereof with a test compound and determining the ability of the test
20 compound to bind to LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 protein or biologically active portion thereof. Binding of the test compound to LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 protein can be determined either directly or indirectly using the approaches described above. In a
25 preferred embodiment, the assay includes contacting LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 protein or biologically active portion thereof with a known compound which binds LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 to form an assay mixture, contacting the assay mixture with
30 a test compound, and determining the ability of the test compound to interact with LIG46, LIG56, Tgtp, LRG-47, RC10-

II, or Stral3 protein, wherein determining the ability of the test compound to interact with LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stral3 protein comprises determining the ability of the test compound to preferentially bind to
5 LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stral3 or biologically active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-free assay comprising contacting LIG46 or LIG56 protein or biologically
10 active portion thereof with a test compound and determining the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stral3 protein or a biologically active portion thereof. Determining the ability of the test
15 compound to modulate the activity of LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stral3 can be accomplished, for example, by determining the ability of LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stral3 protein to bind to a LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stral3 by one of the methods described
20 herein for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stral3 can be accomplished by determining the ability of the agent to alter the activity of LIG46, LIG56, Tgtp,
25 LRG-47, RC10-II, or Stral3 target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined.

In yet another embodiment, the cell-free assay comprises contacting the LIG46, LIG56, Tgtp, LRG-47, RC10-
30 II, or Stral3 protein or biologically active portion thereof with a known compound which binds LIG46, LIG56, Tgtp, LRG-

47, RC10-II, or Stral3 to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stral3 protein, wherein
5 determining the ability of the test compound to interact with a LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stral3 protein comprises determining the ability of the LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stral3 protein to preferentially bind to or modulate the activity of a LIG46,
10 LIG56, Tgtp, LRG-47, RC10-II, or Stral3 target molecule.

For membrane-bound proteins such as LIG46, in one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of LIG46 protein, or a biologically active portion thereof, on the cell surface is
15 contacted with a test compound and the ability of the test compound to bind to a LIG46 protein determined. The cell, for example, can be a yeast cell or a cell of mammalian origin. Determining the ability of the test compound to bind to the LIG46 protein can be accomplished, for example,
20 by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the LIG46 protein or biologically active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with
25 ^{125}I , ^{35}S , ^{14}C , or ^3H , either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, test compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the
30 enzymatic label detected by determination of conversion of an appropriate substrate to product. In a preferred

embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of LIG46 protein, or a biologically active portion thereof, on the cell surface with a known compound which binds LIG46 to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a LIG46 protein, wherein determining the ability of the test compound to interact with a LIG46 protein comprises determining the ability of the test compound to preferentially bind to LIG46 or a biologically active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of LIG46 protein, or a biologically active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of the LIG46 protein or biologically active portion thereof. Determining the ability of the test compound to modulate the activity of LIG46 or a biologically active portion thereof can be accomplished, for example, by determining the ability of the LIG46 protein to bind to or interact with a LIG46 target molecule. As used herein, a "target molecule" is a molecule with which a LIG46 protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses a LIG46 protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. A LIG46 target molecule can be a non-LIG46 molecule or a LIG46 protein or polypeptide of the present invention. In one embodiment, a LIG46 target

molecule is a component of a signal transduction pathway which facilitates transduction of an extracellular signal (e.g., a signal generated by binding of a compound to a membrane-bound LIG46 molecule) through the cell membrane and
5 into the cell. The target, for example, can be a second intercellular protein which has catalytic activity or a protein which facilitates the association of downstream signaling molecules with LIG46.

Determining the ability of the membrane bound LIG46
10 protein to bind to or interact with a LIG46 target molecule can be accomplished by one of the methods described above for determining direct binding. In a preferred embodiment, determining the ability of the LIG46 protein to bind to or interact with a LIG46 target molecule can be accomplished by
15 determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting catalytic/enzymatic activity or detecting a cellular response.

The cell-free assays of the present invention are
20 amenable to use of both the soluble form or the membrane-bound form of LIG46. In the case of cell-free assays comprising the membrane-bound form of LIG46, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of LIG46 is maintained in solution.
25 Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)n, 3-[(3-
30 cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-

propane sulfonate (CHAPSO), or N-dodecyl=N,N-dimethyl-3-ammonio-1-propane sulfonate.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stral3 or the corresponding target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stral3, or interaction of LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stral3 with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical; St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stral3 protein, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtitre plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes

can be dissociated from the matrix, and the level of LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on
5 matrices can also be used in the screening assays of the invention. For example, either LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 or the corresponding target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated LIG46, LIG56, Tgtp, LRG-47,
10 RC10-II, or Stra13 or the corresponding target molecule can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (e.g., biotinylation kit, Pierce Chemicals; Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce
15 Chemical). Alternatively, antibodies reactive with LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 or the corresponding target molecule but which do not interfere with binding of the LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 protein to its target molecule can be derivatized to the wells of
20 the plate, and unbound target or LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using
25 antibodies reactive with the LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 or corresponding target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 or corresponding target molecule.
30 In another embodiment, modulators of LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stra13 expression are identified

in a cell-based assay in which a cell is contacted with a candidate compound and the expression of LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stral3 mRNA or protein in the cell is determined. The level of expression of LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stral3 mRNA or protein in the presence of the candidate compound is compared to the level of expression of LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stral3 mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stral3 expression based on this comparison. For example, when expression of LIG46 mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of LIG46 mRNA or protein expression. Alternatively, when expression of LIG46 mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of LIG46 mRNA or protein expression. The level of LIG46 mRNA or protein expression in the cells can be determined by methods described herein for detecting LIG46 mRNA or protein.

In another embodiment, modulators of LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stral3 activity are identified in a cell-based assay in which a cell is contacted with a candidate compound and the activity of LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stral3 mRNA or protein in the cell is determined. The level of activity of LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stral3 mRNA or protein in the presence of the candidate compound is compared to the level of activity of LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stral3

mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stral3 activity based on this comparison.

5 For example, when activity of LIG46 is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of LIG46 mRNA or protein expression. Alternatively, when the activity of
10 LIG46 is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of LIG46 activity.

 In yet another aspect of the invention, LIG46,
15 LIG56, Tgtp, LRG-47, RC10-II, or Stral3 protein can be used a "bait protein" in a two-hybrid assay or three hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos et al. (1993) *Cell* 72:223-232; Madura et al. (1993) *J. Biol. Chem.* 268:12046-12054; Bartel et al. (1993) *Bio/Techniques* 14:920-
20 924; Iwabuchi et al. (1993) *Oncogene* 8:1693-1696; and PCT Publication No. WO 94/10300), to identify other proteins, which bind to or interact with LIG46, LIG56, Tgtp, LRG-47, RC10-II, or Stral3 and modulate activity.

 The two-hybrid system is based on the modular nature
25 of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for the protein of interest, e.g., Stral3, is fused to a gene encoding the DNA binding domain
30 of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences,

that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with Stra13.

This invention further pertains to novel agents identified by the above-described screening assays and uses thereof for treatments as described herein.

B. Detection Assays

Portions or fragments of the cDNA LIG46 and LIG56 sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

1. Chromosome Mapping

Once the sequence (or a portion of the sequence) of

a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. Accordingly, LIG46 or LIG56 nucleic acid molecules described herein or fragments thereof, can be used to map the location of LIG46 or LIG56 genes on a chromosome. The mapping of the LIG46 or LIG56 sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, LIG46 or LIG56 genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the LIG46 or LIG56 sequences. Computer analysis of LIG46 or LIG56 sequences can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the LIG46 or LIG56 sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (e.g., human and mouse cells).

As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but human cells can, the one human chromosome that contains the gene encoding the needed enzyme, will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human

chromosomes. (D'Eustachio et al. (1983) *Science* 220:919-924). Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

5 PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the LIG46 or LIG56 sequences to design oligonucleotide primers,
10 sublocalization can be achieved with panels of fragments from specific chromosomes. Other mapping strategies which can similarly be used to map a LIG46 or LIG56 sequence to its chromosome include *in situ* hybridization (described in Fan et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:6223-27),
15 pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries.

 Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be
20 used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical like colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa.
25 A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding
30 to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and

more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of this technique, see Verma et al., (Human Chromosomes: A Manual of Basic Techniques (Pergamon Press, New York, 1988)).

5 Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes
10 actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

 Once a sequence has been mapped to a precise
15 chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between
20 genes and disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, e.g., Egeland et al. (1987) *Nature*, 325:783-787.

 Moreover, differences in the DNA sequences between
25 individuals affected and unaffected with a disease associated with the LIG46 or LIG56 gene can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the
30 particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural

alterations in the chromosomes such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence.

Ultimately, complete sequencing of genes from several
5 individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

2. Tissue Typing

The LIG46 or LIG56 sequences of the present
10 invention can also be used to identify individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA
15 is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification.

This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The sequences of the
20 present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

Furthermore, the sequences of the present invention can be used to provide an alternative technique which determines the actual base-by-base DNA sequence of selected
25 portions of an individual's genome. Thus, the LIG46 or LIG56 sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

30 Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique

individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences.

The sequences of the present invention can be used to obtain such identification sequences from individuals and
5 from tissue. The LIG46 or LIG56 sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between
10 individual humans occurs with a frequency of about once per each 500 bases. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the
15 noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:1 can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100
20 bases. If predicted coding sequences, such as those in SEQ ID NO:3 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from LIG46 or LIG56 sequences described herein is used to generate a unique identification
25 database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

3. Use of Partial LIG46 or LIG56 Sequences in

Forensic Biology

DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NO:1 are particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the LIG46 or LIG56 sequences or portions thereof, e.g., fragments derived from the noncoding regions of SEQ ID NO:1 having a length of at least 20 or 30 bases.

The LIG46 or LIG56 sequences described herein can further be used to provide polynucleotide reagents, e.g., labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue, e.g., brain tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such LIG46 or LIG56 probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these reagents, e.g., LIG46 or LIG56 primers or probes can be used to screen tissue culture for contamination (i.e., screen for the presence of a mixture of different types of cells in a culture).

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application are hereby incorporated by reference.

EXAMPLES

Example 1: Identification of Leptin Induced Genes

The leptin induced genes of the invention were identified by comparing the expression pattern of leptin-treated murine neuronal cells expressing OB-RL with the expression pattern of otherwise identical treated cells not expressing OB-RL.

Preparation of Ob Receptor Expressing Neuronal Cells

An adenovirus vector expressing long form murine OB receptor (ObR-L) (Bauman et al. (1996) Proc. Nat'l. Acad. Sci. USA 93:8374-78) was prepared using standard techniques.

A high titer viral stock carrying this vector was prepared and used to infect GT1-7 murine neuronal cells. The infected cells were incubated in standard growth medium for 48 hours and then tested for ObR-L expression by measuring binding of labelled leptin ((1995) Cell 83:1263-71). This assay demonstrated that the infected cells express ObR-L.

Preparation of a Subtracted Library

The ObR-L expressing murine neuronal cells described above were starved for four hours by growth in serum-free medium. A sample of the starved cells was stimulated by incubation in the presence of 200 ng/ml murine leptin for three hours. A second sample of starved cells was mock-stimulated. Total RNA was isolated from both cell samples and used to create cDNA using the SMART PCR™ cDNA synthesis kit (Clontech, Inc.; Palo Alto, CA). The two cDNA pools (generated from total RNA harvested from untreated and leptin-treated cells) created as described above were used to create a subtracted library using the Clontech PCR-Select cDNA Subtraction Kit (Clontech, Inc.).

Screening of the Subtracted Library and Analysis of Positive Clones

The clones in the subtracted library were cloned into T/A vector plasmid T-Adv (Advantage PCR Cloning Kit; Clontech, Inc.). Plasmid specific flanking primers were used to PCR amplify cDNA inserts from the library. The PCR products were then used to create microarrays on nylon filters. The microarrays were probed with labeled cDNA from the subtracted library. Positive clones identified on the the microarray were sequenced, and differential expression of the positive clones was confirmed by virtual Northern analysis on the original treated and untreated samples (pre-

subtracted cDNA generated from from the original cell samples). Additionally, a subset of these clones were analyzed for brain and peripheral tissue distribution by Northern blotting.

5 Two positive clones which appeared to represent novel genes were used to probe a murine whole brain library in order to identify full-length clones. This resulted in the identification of LIG46 and LIG56.

Six of the leptin induced genes identified as
10 described above, LIG46, LIG56, Tgtp, LRG-47, RC10-II, and Stra13 are described in greater detail below.

Example 2: Characterization of Murine and Human LIG46 cDNA and Protein

15 The LIG46 cDNA isolated as described above (SEQ ID NO:1) has a 1191 nucleotide open reading frame (nucleotides ___- ___ of SEQ ID NO:1; SEQ ID NO:3) which encodes a 397 amino acid protein (SEQ ID NO:2). This protein includes a predicted signal sequence of about 32 amino acids (from
20 amino acid 1 to about amino acid 32 of SEQ ID NO:2) and a predicted mature protein of about 365 amino acids (from about amino acid 33 to amino acid 397 of SEQ ID NO:2; SEQ ID NO:4). The extracellular domain of LIG46 extends from about amino acid 33 to about amino acid 302. LIG46 protein
25 possesses one predicted transmembrane domain which extends from about amino acid 303 (extracellular end) to about 320 (intracellular end) of SEQ ID NO:2. The cytoplasmic domain of LIG46 extends from about amino acid 321 to about amino acid 397.

30 LIG46 protein has some sequence similarity to a number of galactosyltransferases. Galactosyltransferases

have been implicated in developmental processes. In addition, galactosyltransferases may play a role in cell to cell signaling by modifying the carbohydrate repertoire on cell surface receptors to activate, inhibit or otherwise
5 modify (e.g., by alter receptor affinity for a ligand) receptor activity. Thus, LIG46 may play a role body weight regulation by influencing cell to cell signaling mediated by molecules involved in body weight regulation, e.g., leptin.

The LIG46 polypeptide sequence of SEQ ID NO:2
10 includes potential N-glycosylation sites at amino acids 30-33, 79-82, 89-92, 127-173, and 219-222; potential protein kinase C phosphorylation sites at amino acids 54-56, 202-204, 221-223, 323-325, and 377-379; potential casein kinase II phosphorylation sites at amino acids 31-34, 94-97, 185-
15 188, 221-224, 234-237, and 368-371; a potential tyrosine kinase phosphorylation site at amino acids 115-122; and a potential amidation site at amino acids 3-6.

Portions of LIG46 are similar to certain galactosyltransferases. Figure 2 depicts a series of
20 alignments of portions of the amino acid sequence of LIG46 with portions of a number of galactosyltransferases, including: *Mus musculus* UDP-Gal: betaGlcNAc beta 1,3-galactosyltransferase-I (Accession Number AF029790; SEQ ID NO:___); *Mus musculus* IPP-Gal: betaGlcNAc beta 1,3-
25 galactosyltransferase-III (Accession Number AF029792); *Drosophila melanogaster* neurogenic secreted signalling protein (Accession Number U41449; SEQ ID NO:___); and *Homo sapiens* UDP-galactose: 2-acetamido-2-deoxy-D-glucose3beta-galactosyltransferase (Accession Number Y15014; SEQ ID
30 NO:___). A majority sequence is depicted above the solid line. Conserved residues are shaded. These residues are

more likely conserved in functional variants of LIG46.

Figure 3 is a hydropathy plot of LIG46. Relative hydrophobicity is shown above the dotted line, and relative hydrophilicity is shown below the dotted line.

5 Figure 7 depicts the cDNA sequence of a full-length human LIG46 clone. Figure 8 depicts the predicted amino acid sequence of human LIG46. The human LIG46 cDNA depicted in Figure 7 (SEQ ID NO:__) has a 1191 nucleotide open reading frame which encodes a 397 amino acid protein (SEQ ID
10 NO:__). This protein includes a predicted signal sequence of about 32 amino acids (from amino acid 1 to about amino acid 32 of SEQ ID NO:__) and a predicted mature protein of about 365 amino acids (from about amino acid 33 to amino acid 397 of SEQ ID NO:__; SEQ ID NO:__). Figure 9 depicts
15 an alignment of the cDNA sequences of human LIG46 (upper sequence) and murine LIG46 (lower sequence). Figure 10 depicts an alignment of the predicted amino acid sequences of human LIG46 (upper sequence) and murine LIG46 (lower sequence).

20 Genomic Mapping of LIG46

LIG46 was mapped to human chromosome 2, 17.9 cR₃₀₀₀ telomeric to the Whitehead Institute framework marker D2S290 (LOD score = 15.5) and 23.5 cR₃₀₀₀ centromeric of the
25 Whitehead framework marker WI-6130 (LOD score = 13.6). This region corresponds to cytogenic location 2p12-13, within or just outside the minimal interval for Alström syndrome (Macari et al. (1998) Human Genet. 103:658-61). Alström syndrome is an autosomal recessive disorder characterized by
30 childhood obesity, retinal pigment degeneration, neurogenic deafness, non-insulin dependent diabetes mellitus, chronic

nephropathy, and hyperlipidemia. Other symptoms include: cardiomyopathy, acanthosis nigricans, hypothyroidism, growth hormone deficiency, progressive baldness, hyperuricemia, gynecomastia, and reduced fertility (Russell-Eggitt et al. (1998) Ophthalmology 105:1274-80).

Briefly, the LIG46 gene was mapped using the Genebridge 4 Radiation Hybrid Panel. A pair of primers within the 3' untranslated region of LIG46 (forward-CCATGTTGGGGTCTCACATTAGAG, SEQ ID NO:___; and reverse-GGTAAGTCAGACCAATATCCTGCC, SEQ ID NO:___) were used to amplify DNA from the Genebridge 4 panel. The PCR products were run on a 2% agarose gel, stained with SYBR Gold and scanned. Linkage analysis was performed using the Map Manager QT623 software package.

LIG46 nucleic acid molecules can be used in the diagnosis of Alström syndrome. Moreover, it is possible that mutations in LIG46 cause Alström syndrome. If so, LIG46 polypeptide and nucleic acid molecules as well as antibodies directed against LIG46 and modulators of LIG46 expression or activity can be used to treat Alström syndrome and/or various symptoms of Alström syndrome.

Example 3: Distribution of LIG46 mRNA

The expression of LIG46 in murine tissue was analyzed using Northern blot hybridization. Analysis of total tissue blots revealed that LIG46 is expressed at the highest level in heart and liver followed by lung and kidney, then brain, then spleen testis, and skeletal muscle.

Analysis of LIG46 expression in murine brain revealed that LIG46 is expressed at least in the hypothalamus (including: the arcuate nucleus, the ventral/medial hypothalamus, and

the superchiasmatic nucleus, the hippocampus, the cortex, and the striatum.

Example 4: Secretion of LIG46

LIG46 protein is homologous to *D. melanogaster* brainiac (Goode et al., (1996) Development 122:3863-79), a secreted protein (Fig. 2). As discussed above, LIG46 has a predicted signal sequence at its amino terminus.

Therefore, to determine whether LIG46 protein is secreted, full-length LIG46 (amino acids 1-397) was fused to alkaline phosphatase using methods similar to those previously described (fusion at carboxy-terminus of LIG46; Cheng and Flanagan (1994) Cell 79:157-168; Tartaglia et al. (1995) Cell 83:1263-71). This construct was transiently transfected into human 293T cells.

At 48 hrs post transfection, the growth media was assayed for alkaline phosphatase activity (White et al., (1997) Proc. Natl. Acad. Sci USA 94:10657-10662) using the Great EscAPE alkaline phosphatase detection kit (Clontech, Inc.). A large increase in alkaline phosphatase activity was observed in the growth medium from transfected cells compared to mock tranfected cells, indicating that LIG46 protein is secreted and that the signal sequence of LIG46 is functional.

Example 5: LIG46 Expression is Induced by Leptin *in vivo*

C57BL6 *ob/ob* mice were injected (via the interperitoneum (IP)) with 100 µl of either phosphate buffered saline (PBS) (sham injected) or PBS supplemented with 100 µg leptin (leptin injected) (R&D Systems Inc., Minneapolis, MN). Following a 1 or 3 hr treatment, the

animals were euthanized by CO₂ asphyxiation, the brains were harvested, sliced, and the hypothalamus analyzed by *in situ* hybridization using a 386 base pair radiolabeled antisense probe to the coding region of LIG46. Comparative analysis of hypothalamic slices from sham injected and leptin injected animals indicates that LIG46 transcript is induced in the arcuate nucleus and the ventromedial hypothalamus by leptin.

10 Example 6: The Effect of LIG46 Antisense

Oligodeoxynucleotides on Feeding of Obese (ob/ob) Male Mice

For this study, a phosphothioate-protected antisense oligodeoxynucleotide and its respective control sequence (sense) were synthesized. The antisense

15 oligodeoxynucleotide targets the LIG46 start codon mRNA at position 39.

Antisense: 5' CTT CGA CGC CCC ACA CTC AT 3' (SEQ ID NO:_)

Sense: 5' ATG AGT GTG GGG CGT CGA AG 3' (SEQ ID NO:_)

Male obese *ob/ob* C57BL/6J (45 g) mice were individually

20 housed in macrolon cages (22±2° C; 12:12 h light/dark cycle with lights off at 6 pm). Tap water and mouse chow diet were given *ad libitum*. Mice were stereotaxically implanted with a chronic guide cannula aimed to the third ventricle (intracerebroventricular) one week prior to this experiment.

25 The effect of LIG46 antisense treatment on leptin-induced decrease in food intake was studied on day 5.

Therefore, mice were treated intracerebroventricularly on days 1 and 3 with 18 µg LIG46 antisense oligodeoxyribonucleotide, 18 µg sense (control)

30 oligodeoxyribonucleotide or 2 µl RNase-free water.

Intracerebroventricular injections were performed at 3 pm

Control and oligodeoxyribonucleotide pre-treatments were followed by an intraperitoneal injection of 1 mg/kg leptin or phosphate-buffered saline (vehicle), performed at 5 pm on day 5 and food intake was measured each four hour after leptin or vehicle application. The results of this study are shown in Fig. 6. The leptin-induced decrease in food intake was far greater in the presence of LIG46 antisense oligonucleotide than LIG46 sense nucleotide or PBS control.

10 Example 7: The Effect of LIG46 Antisense

Oligodeoxynucleotides on Feeding of Lean Male Mice

For this study, a phosphothioate-protected antisense oligodeoxynucleotide and its respective control sequence (sense) were synthesized. The antisense

15 oligodeoxynucleotide targets the LIG46 start codon mRNA at position 39.

Antisense: 5' CTT CGA CGC CCC ACA CTC AT 3' (SEQ ID NO:_)

Sense: 5' ATG AGT GTG GGG CGT CGA AG 3' (SEQ ID NO:_)

Male lean C57BL/6J (24 g) mice were individually housed in macrolon cages (22±2° C; 12:12 h light/dark cycle with lights off at 6 pm). Tap water and mouse chow diet were given *ad libitum*. Mice were stereotaxically implanted with a chronic guide cannula aimed to the third ventricle (intracerebroventricular) one week prior to this experiment.

25 The effect of LIG46 antisense treatment on leptin-induced decrease in food intake was studied on day 5.

Therefore, mice were treated intracerebroventricularly on days 1 and 3 with 18 µg LIG46 antisense oligodeoxyribonucleotide, 18 µg sense (control)

30 oligodeoxyribonucleotide or 2 µl RNase-free water.

Intracerebroventricular injections were performed at 3 pm

Control and oligodeoxyribonucleotide pre-treatments were followed by an intraperitoneal injection of 1 mg/kg leptin or phosphate-buffered saline (vehicle), performed at 5 pm on day 5 and food intake was measured each four hour after leptin or vehicle application. The results of this study are shown in Fig. 11. The LIG46 antisense-induced decrease in food intake was far greater in the presence of leptin than PBS control. Thus, food intake can be decreased in lean mice by decreasing LIG46 protein expression. Moreover, this decrease in food intake is increased when leptin is administered, demonstrating that leptin can sensitize lean mice to the effects of a LIG46 antagonist.

Example 8: Characterization of LIG56 cDNA and Protein

The full-length LIG56 cDNA isolated as described above (SEQ ID NO:5) is shown in Figure 4. This cDNA has a 1200 nucleotide open reading frame (nucleotides 1 - 1200 of SEQ ID NO:5; SEQ ID NO:7) which encodes a 400 amino acid protein (SEQ ID NO:6).

The LIG56 polypeptide sequence of SEQ ID NO:6 includes potential N-glycosylation sites at amino acids 252-255; potential protein kinase C phosphorylation sites at amino acids 67-69, 75-77, 203-205, 218-220, 295-297, and 299-301; potential casein kinase II phosphorylation sites at amino acids 126-129, 170-173, 203-206, 256-259, 291-294, 341-344, and 345-349; a potential tyrosine kinase phosphorylation site at amino acids 233-241; potential N-myristylation sites at amino acids 66-71, 85-90, 116-121, and 308-313; and a potential amidation site at amino acids 63-70.

LIG56 may be a GTP-binding protein. Portions of

LIG56 protein are to similar to one or more murine GTP-binding proteins (Genbank Accession Numbers: L38444; U15636; M63630; U19119; and U53219).

LIG56 protein possesses a GTP-binding protein-like domain (amino acids 12 to 283 of SEQ ID NO:6) and an LRG-47-like domain (amino acids 24-177 of SEQ ID NO:6).

Figure 5 is a hydropathy plot of LIG56. Relative hydrophobicity is shown above the dotted line, and relative hydrophilicity is shown below the dotted line.

Example 9: Distribution of LIG56 mRNA

The expression of LIG56 in murine tissue was analyzed using Northern blot hybridization. Analysis of total tissue blots revealed that LIG56 is expressed at the highest level in heart followed by liver, then kidney, then lung, then skeletal muscle, then spleen. Analysis of LIG56 expression in murine brain revealed that LIG56 is expressed at least in the hippocampus (including, at least, the dentate gyrus).

Example 10: Characterization and mRNA Distribution of Clone 50 (Tgtp)

Sequence analysis of clone 50 identified in the microarray described above revealed that the clone encodes murine Tgtp (Genbank Accession Number L38444), a T cell-specific guanine nucleotide triphosphate-binding protein (Carlow et al. (1994) *J. Immunol.* 154:1724-34).

The expression of clone 50 in murine tissue was analyzed using Northern blot hybridization. Analysis of total tissue blots revealed that clone 50 is expressed at the highest level in heart followed by kidney, then lung and

skeletal muscle, then liver. Analysis of clone 50 expression in murine brain revealed that clone 50 is expressed at least in the choroid plexus.

5 Example 11: Characterization and mRNA Distribution of Clone
44 (LRG-47)

Sequence analysis of clone 44 identified in the microarray described above revealed that the clone encodes murine LRG-47 (Genbank Accession Number U19119), a protein
10 that is induced by LPS, IFN- γ , and IFN- α/β and has some homology to GTP-binding proteins (Sorace et al. (1995) *J. Leukocyte Biol.* 58:477-84).

The expression of clone LRG-47 mRNA in murine tissue was analyzed using Northern blot hybridization. Analysis of
15 total tissue blots revealed that LRG-47 is expressed at the highest level in heart followed by kidney, then liver, then skeletal muscle, then lung, then spleen. Analysis of LRG-47 mRNA expression in murine brain revealed that LRG-47 is expressed at least in the cortex, the hippocampus, the
20 choroid plexus, the medial habenular nucleus, and the hypothalamus (including at least: the arcuate nucleus and the paraventricular nucleus).

Example 12: LRG-47 is Induced by Leptin *in vivo*

25 C57BL6 *ob/ob* mice were injected (IP) with 100 μ l of either PBS or PBS supplemented with 100 μ g leptin (R&D Systems Inc.). After 60 min, the animals were euthanized by CO₂ asphyxiation, the brains harvested, sliced, and the hypothalamus analyzed by *in-situ* hybridization using a 762
30 base pair radiolabeled antisense probe against sequences in the 5' untranslated region of the LRG-47 transcript.

Comparative analysis of hypothalamic slices from sham
injecting and leptin injected animals indicates that the
LRG-47 transcript is induced in the arcuate nucleus by
leptin, demonstrating that the LRG-47 transcript is a
5 leptin-induced gene *in vivo*.

Example 13: Characterization and mRNA Distribution of Clone 10 (RC10-11)

Sequence analysis of clone 10 identified in the
10 microarray described above revealed that the clone encodes
murine RC10-II (Genbank Accession Number D21800), a subunit
of the 20S proteasome of rat embryonic brain (Nishimura et
al. (1993) *FEBS Lett.* 336:462-66). It has been suggested
that RC10-II is a proteasomal subunit that is required for
15 expression of tryptic activity (Nishimura et al., *supra*).

The expression of clone 10 mRNA in murine tissue was
analyzed using Northern blot hybridization. Analysis of
total tissue blots revealed that clone 10 is expressed at
the highest level in heart, liver, skeletal muscle, and
20 kidney, followed by brain, lung, and testis. Analysis of
clone 10 mRNA expression in murine brain revealed that clone
10 is expressed at least in the cortex, hippocampus,
habenular nucleus, thalamus, and hypothalamus (including the
arcuate nucleus and ventromedial hypothalamus).

Example 14: Characterization and mRNA Distribution of Clone 67 (Stra13)

Sequence analysis of clone 67 identified in the
microarray described above revealed that the clone encodes
30 Stra13 (Genbank Accession Number AF010305), retinoic acid-
inducible helix-loop-helix protein (Boudjelal et al. (1997)

Genes Dev. 11:2052-65). Stra13 may act as a repressor of activated transcription and is thought to play a role in neuronal differentiation (Boudjelal et al., *supra*).

The expression of clone 67 mRNA in murine tissue was analyzed using Northern blot hybridization. Analysis of total tissue blots revealed that clone 10 is expressed at the highest level in liver followed by heart, then skeletal muscle, then brain, then kidney. Analysis of clone 67 mRNA expression in murine brain revealed that clone 67 is expressed at least in the cortex, hippocampus (CA1, CA2 and dentate gyrus), lateral thalamus, hypothalamus (arcuate nucleus).

Example 15: Stra13 is Induced by Leptin *in vivo*

C57BL6 *ob/ob* mice were injected IP with 100 μ l of either PBS or PBS supplemented with 100 μ g leptin (R&D Systems, Inc.). After 60 min, the animals were euthanized by CO₂ asphyxiation, the brains harvested, sliced, and the hypothalamus analyzed by *in situ* hybridization using 328 base pair radiolabeled antisense probe against sequences in the 5' untranslated region of the LRG-47 transcript. Comparative analysis of hypothalamic slices from sham injected and leptin injected animals indicates that the Stra13 transcript is induced in the arcuate nucleus by leptin, supporting the claim that the Stra13 transcript is a leptin-induced gene *in vivo*.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the
5 invention described herein. Such equivalents are intended to be encompassed by the following claims.

What is claimed is: